

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 10, 2004, 10:48:44 ; Search time 30.8071 Seconds  
(without alignments)  
1457.493 Million cell updates/sec

Title: US-10-067-385-8\_COPY\_600\_773

Perfect score: 897  
Sequence: 1 KIVKDFARNYTKVFKFILNK.....ATVLDKNINISSKSTNNPNK 174

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

SPTREMBL\_23.\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_virus:\*  
16: sp\_bacteriaph:\*  
17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	897	100.0	2119	2	Q9AHT5
2	897	100.0	2140	16	Q9AHT5
3	894	99.7	2144	16	Q9AHT5
4	893	99.6	2144	2	Q9AHT5
5	119	13.3	775	16	Q8CPK8
6	114.5	12.8	1038	13	Q90784
7	114	12.7	609	5	Q812K8
8	112.5	12.5	346	5	Q90UG0
9	112.5	12.5	354	5	Q25995
10	112.5	12.5	354	5	Q81U55
11	112.5	12.5	361	5	Q95P15
12	112.5	12.5	379	5	Q9U6C4
13	112	12.5	829	5	Q81SF3
14	111.5	12.4	379	5	Q25706
15	110.5	12.3	379	5	Q25705
16	110	12.3	380	5	Q26019

17	109.5	12.2	3008	5	Q81436
18	109	12.2	951	5	Q96229
19	108.5	12.1	600	5	Q77355
20	107.5	12.0	470	10	Q9FUK9
21	107.5	12.0	948	5	Q9U4U6
22	106	11.8	3026	5	Q81LS9
23	105.5	11.8	329	5	Q9NFV9
24	105.5	11.8	2081	10	Q9LH98
25	105	11.7	3127	5	Q81DA0
26	104	11.6	540	10	Q9AC59
27	104	11.6	540	10	Q9SA84
28	103.5	11.5	325	5	Q44016
29	103	11.5	238	5	Q812Z6
30	102.5	11.4	385	5	Q93424
31	102.5	11.4	449	5	Q81HW3
32	102	11.4	622	16	Q96Q41
33	102	11.4	1419	5	Q9UBD4
34	102	11.4	1612	5	Q81FNO
35	101.5	11.3	382	16	Q9PPL5
36	101.5	11.3	382	5	Q9V7J0
37	101.5	11.3	535	5	Q17595
38	101.5	11.3	556	5	Q9V7I9
39	101.5	11.3	785	5	Q9GQ82
40	101.5	11.3	1377	5	Q81596
41	101.5	11.3	1434	5	Q81J77
42	101	11.3	700	3	Q07457
43	101	11.2	1490	5	Q81416
44	100.5	11.2	354	2	Q93CL2
45	100.5	11.2	1373	5	Q81353

## ALIGNMENTS

RESULT 1  
ID Q9AHT5 PRELIMINARY; PRT; 2119 AA.  
AC Q9AHT5;  
DT 01-JUN-2001 (TRENBLREL. 17, Created)  
DT 01-JUN-2001 (TRENBLREL. 17, Last sequence update)  
DT 01-MAR-2003 (TRENBLREL. 23, Last annotation update)  
DB Serine protease (Fragment).  
DR PRTA.  
OS Streptococcus pneumoniae.  
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_Taxid=1313;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=N4;  
RX MEDLINE=21116976; PubMed=11179332;  
RA Witzmann T.M., Heinrichs J.H., Adamou J.E., Erwin A.L., Kunsch C., Choi G.H., Barash S.C., Rosen C.A., Masure H.R., Tuomanen E., Gayle A., Brewah Y.A., Walsh W., Barren P., Lathigra R., Hanson M., Langermann S., Johnson S., Koenig S.;  
RA "Use of a Whole Genome Approach to Identify Vaccine Molecules Affording Protection against Streptococcus pneumoniae Infection." Infect. Immun. 69:1593-1598(2001).  
RT - SUBCELLULAR LOCATION: ATTACHED TO THE CELL WALL PEPTIDOGLYCAN BY AN AMIDE BOND (BY SIMILARITY).  
CC EMBL: AF291699; AKK159.1; -.  
CC HSRP; P00782; 2SBR.  
CC MEROPS; S08.064; -.  
DR InterPro: IPR001899; Gram\_pos\_anchor.  
DR InterPro: IPR006192; LPXTG.  
DR InterPro: IPR003137; PA.  
DR InterPro: IPR000209; Peptidase\_S8.  
DR InterPro: IPR001680; WD40.  
DR Pfam; PR00746; Gram\_pos\_anchor; 1.  
DR Pfam; PR02225; PA; 1.  
DR Pfam; PR00082; Peptidase\_S8; 2.  
DR PRINTS; PR00723; SUBTILISIN.  
DR TIGRfam; TIGR01167; LPXTG\_anchor; 1.

DR PROSITE; PS50847; GRAM\_POS\_ANCHORING; 1.  
 DR PROSITE; PS50840; PA; 1.  
 DR PROSITE; PS00137; SUBTILASE\_HIS; 1.  
 DR PROSITE; PS00138; SUBTILASE\_SER; 1.  
 DR PROSITE; PS00678; WD\_REPEATS\_1; 1.  
 FT Cell wall; peptidoglycan-anchor; Protease.  
 FT NON\_TER 1  
 SQ SEQUENCE 2119 AA; 238226 MW; 517P9B7F6B960A6A CRC64;

Query Match 100.0%; Score 897; DB 2; Length 2119;  
 Best Local Similarity 100.0%; Pred. No. 7, 7e-54;  
 Matches 174; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KIVVKDPARTVTKPEFLNKDGTGSEVSELKPHRYVTIIONGKMSSTIVSEDFILPYKKG 60  
 DB 1912 KIVVKDPARTVTKPEFLNKDGTGSEVSELKPHRYVTIIONGKMSSTIVSEDFILPYKKG 1971  
 QY 61 ELEKGYQFDGMEISGEGKKGADGYVNLSDOTFIKPFYFKIIEKKEEENKPTDVSXKGD 120  
 DB 1972 ELEKGYQFDGMEISGEGKKGADGYVNLSDOTFIKPFYFKIIEKKEEENKPTDVSXKGD 2031

QY 121 NPQVNSQLNESHKEDLQREHHSQKSDSTKDVATATVLDKNNTSSKSTNNPNK 174  
 DB 2032 NPQVNSQLNESHKEDLQREHHSQKSDSTKDVATATVLDKNNTSSKSTNNPNK 2085

RESULT 2  
 Q97RY6 PRELIMINARY; PRT; 2140 AA.  
 ID Q97RY6;  
 AC Q97RY6;  
 DT 01-OCT-2001 (TrEMBLrel. 18, Created)  
 DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE Serine protease, subtilase family.  
 GN SP0641.  
 OS Streptococcus pneumoniae.  
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 OC Streptococcus.  
 OX NCBI\_TaxID=1313;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=TI6R4;  
 RX MEDLINE=21357209; PubMed=11463916;  
 RA Tetteilin H., Nelson K.E., Paulsen I.T., Eissen J.A., Read T.D.,  
 RA Peterson S., Heidelberg J., Deboy R.T., Haft D.H., Dodson R.J.,  
 RA Durkin A.S., Grimm M., Kolonay J.F., Nelson W.C., Peterson J.D.,  
 RA Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,  
 RA Holtzapple E., Khouri H., Wolf A.M., Utterback T.R., Hansen C.L.,  
 RA McDonald L.A., Feldblyum T.V., Angiuoli S., Dickinson T., Hickey E.K.,  
 RA Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,  
 RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;  
 RT "Complete genome sequence of a virulent isolate of Streptococcus pneumoniae.";  
 RL Science 293:498-506(2001).  
 DR EMBL; AE007373; AAK74791.1; -.  
 DR MEROPS; S08\_064; -.  
 DR TIGR; SP0641; -.  
 DR InterPro; IPR001839; Gram\_pos\_anchor.  
 DR InterPro; IPR006192; LPXTG.  
 DR InterPro; IPR003137; PA.  
 DR InterPro; IPR000209; Peptidase\_S8.  
 DR InterPro; IPR001680; WD40.  
 DR Pfam; PF00746; Gram\_pos\_anchor; 1.  
 DR Pfam; PF02225; PA; 1.  
 DR Pfam; PF00082; Peptidase\_S8; 2.  
 DR PRINTS; PR00723; SUBTILASIN.  
 DR TIGRPFAM; TIGR01167; LPXTG\_anchor; 1.  
 DR PROSITE; PS50847; GRAM\_POS\_ANCHORING; 1.  
 DR PROSITE; PS50840; PA; 1.  
 DR PROSITE; PS00137; SUBTILASE\_HIS; 1.  
 DR PROSITE; PS00138; SUBTILASE\_SER; 1.  
 DR PROSITE; PS00678; WD\_REPEATS\_1; 1.  
 DR PROSITE; PS00678; WD\_REPEATS\_1; 1.  
 KW Protease; Complete proteome.

SQ SEQUENCE 2140 AA; 240426 MW; FA44AD8E2938B334 CRC64;

Query Match 100.0%; Score 897; DB 16; Length 2140;  
 Best Local Similarity 100.0%; Pred. No. 7, 8e-54;  
 Matches 174; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KIVVKDPARTVTKPEFLNKDGTGSEVSELKPHRYVTIIONGKMSSTIVSEDFILPYKKG 60  
 DB 1933 KIVVKDPARTVTKPEFLNKDGTGSEVSELKPHRYVTIIONGKMSSTIVSEDFILPYKKG 1992  
 QY 61 ELEKGYQFDGMEISGEGKKGADGYVNLSDOTFIKPFYFKIIEKKEEENKPTDVSXKGD 120  
 DB 1993 ELEKGYQFDGMEISGEGKKGADGYVNLSDOTFIKPFYFKIIEKKEEENKPTDVSXKGD 2052  
 QY 121 NPQVNSQLNESHKEDLQREHHSQKSDSTKDVATATVLDKNNTSSKSTNNPNK 174  
 DB 2053 NPQVNSQLNESHKEDLQREHHSQKSDSTKDVATATVLDKNNTSSKSTNNPNK 2106

RESULT 3  
 O8DQP7 PRELIMINARY; PRT; 2144 AA.  
 ID O8DQP7;  
 AC O8DQP7;  
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE Cell wall-associated serine proteinase PrtA (EC 3.4.21.-).  
 GN PrtA OR SP0561.  
 GN Streptococcus pneumoniae (strain ATCC BAA-255 / R6).  
 OS Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 OC Streptococcus.  
 OX NCBI\_TaxID=171101;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21429245; PubMed=11544234;  
 RA Hoskins J., Alborn W.R. Jr., Arnold J., Blaszczyk L.C., Burgess S.,  
 RA DeHoff B.S., Betrem S.T., Fritz L., Fu D.-J., Fuller W., Geringer C.,  
 RA Gilmore R., Glass J.S., Khoja H., Kraft A.R., Lagace R.B.,  
 RA LeBlanc D.J., Lee L.N., Leikowitz B.J., Lu J., Matsushima P.,  
 RA McAhren S.M., McHenry M., McLeaster K., Mundy C.W., Niclas T.I.,  
 RA Norris F.H., O'Gara M., Peery R.B., Robertson G.T., Rocky P.,  
 RA Sun P.-M., Winkler M.E., Yang Y., Young-Bellido M., Zhao G.,  
 RA Zook C.A., Baltz R.H., Jaskunas S.R., Kosteck P.R. Jr., Skatrud P.L.,  
 RA "Genome of the bacterium Streptococcus pneumoniae strain R6.";  
 RT J. Bacteriol. 183:5709-5717(2001).  
 DR EMBL; AE008434; AAK9365.1; -.  
 KW Hydrolyase; Complete proteome.  
 SQ SEQUENCE 2144 AA; 240436 MW; 8C1B4B1DBC503A0C CRC64;

Query Match 99.7%; Score 894; DB 16; Length 2144;  
 Best Local Similarity 99.4%; Pred. No. 1, 3e-53;  
 Matches 173; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 KIVVKDPARTVTKPEFLNKDGTGSEVSELKPHRYVTIIONGKMSSTIVSEDFILPYKKG 60  
 DB 1937 KIVVKDPARTVTKPEFLNKDGTGSEVSELKPHRYVTIIONGKMSSTIVSEDFILPYKKG 1996  
 QY 61 ELEKGYQFDGMEISGEGKKGADGYVNLSDOTFIKPFYFKIIEKKEEENKPTDVSXKGD 120  
 DB 1997 ELEKGYQFDGMEISGEGKKGADGYVNLSDOTFIKPFYFKIIEKKEEENKPTDVSXKGD 2056  
 QY 121 NPQVNSQLNESHKEDLQREHHSQKSDSTKDVATATVLDKNNTSSKSTNNPNK 174  
 DB 2057 NPQVNSQLNESHKEDLQREHHSQKSDSTKDVATATVLDKNNTSSKSTNNPNK 2110

RESULT 4  
 O9S4M8 PRELIMINARY; PRT; 2144 AA.  
 ID O9S4M8;  
 AC O9S4M8;  
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)

DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)  
 DE Cell wall-associated serine proteinase Prta precursor.  
 GN PRTA.  
 OS Streptococcus pneumoniae.  
 CN Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 OC Streptococcus.  
 CC NCBI\_TaxID=1313;  
 RX NCBI\_TaxID=1313;  
 RN SEQUENCE FROM N.A.  
 RP STRAIN=3.B;  
 RC Behe G., ten Thoren E., Bongers R.J.M., Heinz H.-P., Zysk G.;  
 RT "Cloning and sequencing of a novel surface protease of Streptococcus pneumoniae."  
 RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.  
 CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE CELL WALL PEPTIDOGLYCAN BY AN AMIDE BOND (BY SIMILARITY).  
 CC EMBL: AF127143; AAD48399.1; -.  
 DR HSP; P00782; 2SST.  
 DR MEROPS; S08.064; -.  
 DR InterPro: IPR001899; Gram\_pos\_anchor.  
 DR InterPro: IPR006192; LPTG.  
 DR InterPro: IPR003137; PA.  
 DR InterPro: IPR000209; Peptidase\_S8.  
 DR InterPro: IPR001680; WD40.  
 DR Pfam: PF00746; Gram\_pos\_anchor; 1.  
 DR Pfam: PF02225; PA; 1.  
 DR Pfam; PF00082; Peptidase\_S8; 2.  
 DR PRINTS; PR00723; SUBTILISIN.  
 DR TIGRfams; TIGR01167; LPTG\_anchor; 1.  
 DR PROSITE; PS08447; GRAM\_POS\_ANCHORING; 1.  
 DR PROSITE; PS00137; SUBTILASE\_HIS; 1.  
 DR PROSITE; PS00138; SUBTILASE\_SER; 1.  
 DR PROSITE; PS00678; WD\_REPEATS\_1; 1.  
 KW Cell wall; Peptidoglycan-anchor; Signal.  
 FT SIGNAL 1 19  
 FT CHAIN 20 2144  
 FT FT CELL WALL-ASSOCIATED SERINE PROTEINASE  
 FT PRTA.  
 SQ SEQUENCE 2144 AA; 240724 MW; 2052511470741331 CRC64;  
 Query Match 99.6%; Score 893; DB 2; Length 2144;  
 Best Local Similarity 98.9%; Pred. No. 1.5e-53;  
 Matches 172; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 KIVKDFARNITVKEFLNKOTGEVSEIKPHRVTVITTONKGMSTIYSEEDFILPVYKG 60  
 DB 1937 KIVKDFARNITVKEFLNKOTGEVSEIKPHRVTVITTONKGMSTIYSEEDFILPVYKG 1996  
 QY 61 ELEKGYQDGMWISGFEGKQAGYVYNLSKOTFIKPVFKIEKKEBKPTFDVSKKDD 120  
 DB 1997 ELEKGYQDGMWISGFEGKQAGYVYNLSKOTFIKPVFKIEKKEBKPTFDVSKKDD 2056  
 QY 121 NPQVHNSQLNESHKREDLQREHSQKSDSTKDVATVLDKNNISSKSTNNENK 174  
 DB 2057 NPQVHNSQLNESHKREDLQREHSQKSDSTKDVATVLDKNNISSKSTNNENK 2110  
 RESULT 5  
 Q8CPK8 PRELIMINARY; PRT; 775 AA.  
 ID Q8CPK8  
 AC 08CPK8;  
 DT 01-MAR-2003 (TREMBLrel. 23, Created)  
 DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)  
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)  
 DE Penicillin-binding protein 1.  
 GN SB0856.  
 OS Staphylococcus epidermidis.  
 OC Bacteria; Firmicutes; Bacillales; Staphylococcus.  
 CC NCBI\_TaxID=1282;  
 RX NCBI\_TaxID=1282;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 12228;  
 RA Zhang Y., Ren S., Li H., Fu G., Lu L., Lu G., Jia J., Tu Y., Qin Z.,

RA Chen Z., Wen Y.;  
 RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AB016746; AA004453.1; -.  
 KW Complete proteome.  
 CC SEQUENCE 775 AA; 86354 MW; 4A00563A7BB8777C CRC64;  
 Query Match 13.3%; Score 119; DB 16; Length 775;  
 Best Local Similarity 27.0%; Pred. No. 2;  
 Matches 47; Conservative 24; Mismatches 57; Indels 46; Gaps 9;  
 QY 20 KDTGEVSEIKPHRVTVITTONKGMSTIYSEEDFILPVYK-----GILEKGYQDGMW-- 71  
 DB 614 EDSVNAOSLKP-----ITIGNKQKQIKQSVKSTKYLPHSKWLMPTDGLTMP-DMTGWTX 668  
 QY 72 -EISGF-----GKQAGYVYN--LSKOTFIKPVFKIEKKEBKPTFDVSK-----K 117  
 DB 669 EDVLAFFDLTKIKVSTKNGFVTVNOSISKQLIK-----NKKLEVSLSAED 715  
 QY 118 KKDNPQVHNSQLNESHKREDLQREHSQKSDSTKDVATVLDKNNISSKSTNN 171  
 DB 716 TDDQEKTDSDSDVSKKDKADEDSHNTSSTTKN-----DKSNADSKNDSD 763  
 RESULT 6  
 Q90784 PRELIMINARY; PRT; 1038 AA.  
 ID Q90784  
 AC Q90784;  
 DT 01-NOV-1996 (TREMBLrel. 01, Created)  
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
 DT 01-OCT-2002 (TREMBLrel. 22, Last annotation update)  
 DE Claustirin.  
 OS Gallus gallus (Chicken).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Buteleostomi;  
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
 CC Gallus.  
 RX NCBI\_TaxID=9031;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Brain;  
 RA Cole G.J.;  
 RL Submitted (OCT-1992) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE OF 1-451 FROM N.A.  
 RC TISSUE=Brain;  
 RX MEDLINE=94157526; PubMed=7906711;  
 RA Burg M.A., Cole G.J.;  
 RT "Claustirin, an adhesive neural keratan sulfate proteoglycan, is structurally related to MAP1B."  
 RL J. Neurobiol. 25:1-22(1994).  
 DR EMBL: X67778; CAA47988.1; -.  
 SQ SEQUENCE 1038 AA; 117112 MW; 21306945B510927 CRC64;  
 Query Match 12.8%; Score 114.5; DB 13; Length 1038;  
 Best Local Similarity 23.5%; Pred. No. 5.7;  
 Matches 47; Conservative 29; Mismatches 65; Indels 59; Gaps 6;  
 QY 2 IVVKDFARNITVKEFLNKOTGEVSEIKPHRVTVITTON----- 39  
 DB 514 VTQKDLTGNIASPAVQAKLQRTDSKSLKPAKTTTKQDQKRLKKKHSLSGLVQ 573  
 QY 40 -----GKMSSTIYSEEDFILPVYKGELEKGYQDGMWISGFEGK 79  
 DB 574 LEKPKLSKERTPVKKEKAVPEKTTVAEKDV-----TTKEBQKSTSEKQNS 625  
 QY 80 KDAGYVYNLSKOTFIKPVFK--KIEKKEBKPTFDVSKKQNPQVHNSQLNESHKREDL 138  
 DB 626 EKQDVKPVTKKSKYKKEKAPBEKDKKEPKKEVSKGEKPL-----KKEKPKKDDI 682  
 QY 139 QRE-----EHSQKSDSTKDV 153  
 DB 683 KKEVKKEVKEKKEKKEKKEV 702

```

RESULT 7
ID Q812K8 PRELIMINARY: PRT: 609 AA.
AC Q812K8:
DT 01-MAR-2003 (TRENBLREL. 23, Created)
DT 01-MAR-2003 (TRENBLREL. 23, Last sequence update)
DE Large cyclophilin-like protein.
GN PF11490C.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22255708; PubMed=12368867;
RA Hall N., Pain A., Berriman M., Churcher C., Harris B., Harris D.,
RA Mungall K., Bowman S., Atkin R., Baker S., Barton A., Brooks K.,
RA Bucke C.O., Burrows C., Cherevach I., Chillingworth C.,
RA Chillingworth T., Christodoulou Z., Clark L., Clark R., Corton C.,
RA Crotin A., Davies R., Davis P., Dear P., Dearden F., Doggett J.,
RA Felwell T., Goble A., Goodhead I., Gwilliam R., Hamlin N., Hance Z.,
RA Harper D., Hauser H., Hornsby T., Holroyd S., Horrocks P.,
RA Humphrey S., Jagsels K., James K.D., Johnson D., Kerhornou A.,
RA Knights A., Konfortov B., Kyes S., Larke N., Lawson D., Lennard N.,
RA Line A., Maddison M., Mclean J., Mooney P., Moule S., Murphy L.,
RA Oliver K., Ormond D., Price C., Quail M.A., Rabinowitsch E.,
RA Rajandream M.A., Rutter S., Rutherford K.M., Sanders M., Simmonds M.,
RA Seeger K., Sharp S., Smith R., Squares R., Squares S., Stevens K.,
RA Taylor K., Tivey A., Unwin L., Whitehead S., Woodward J.,
RA Sultson J.E., Craig A., Newbold C., Barrell B.G.;
RT "Sequence of Plasmodium falciparum chromosomes 1, 3-9 and 13."
RL Nature 419:527-531(2002).
DR EMBL; AL29358; CAD51984.1;
SQ SEQUENCE 609 AA; 72552 MW; 8CDF8685FF9A021 CRC64;

Query Match 12.7%; Score 114; DB 5; Length 609;
Best Local Similarity 24.3%; Pred. No. 3.5;
Matches 46; Conservative 38; Mismatches 83; Indels 22; Gaps 6;

QY 2 IVVQDPAFNTYKERTLNKQDGEVSELRKRVTVITQNGKMSSTI---VEEDPFI-- 55
DB 104 ITRKVFNRHTKI---ILNRKTNVYSSSELYK---PHNNENPDPPTTHDKISKEDPIVLQ 157
QY 56 -PVYSGELKGYQFQGWSEISGFEKKDAGYVNLKSDPTFKVFPKIEKKEENKPTD 114
DB 158 DPLNNKTYVASHLNDROEDNEKKKKNIFAPSIQDNGAIMSLKEMEKAKKKEKKEE 217
QY 115 VSKKQDNP---QVNHSQLNESHKREDLQREHESQKSDSTQVATVLD---KNNIS 165
DB 218 EEEBKKKKVYGLQDYNDSNSESSEHSEKCKLKKRQDQYNNKINHSTVDSSENEIKI 277
QY 166 KSTNNPNK 174
DB 278 KCTNYSNPK 286

RESULT 8
ID Q9U0G0 PRELIMINARY: PRT: 346 AA.
AC Q9U0G0:
DT 01-MAY-2000 (TRENBLREL. 13, Created)
DT 01-MAY-2000 (TRENBLREL. 13, Last sequence update)
DT 01-OCT-2002 (TRENBLREL. 22, Last annotation update)
DE Merozoite surface protein 3 (Fragment).
GN MSP3.
OS Plasmodium reichenowi.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5854;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20416497; PubMed=10960178;
RA Okenu D.M.N., Thomas A.W., Conway D.J.;
RT "Allelic lineages of the merozoite surface protein 3 gene in

```

```

RT Plasmodium reichenowi and Plasmodium falciparum."
RL Mol. Biochem. Parasitol. 109:185-188(2000).
DR EMBL; AJ252286; CAB65754.1;
FT NON_TER 1 1
FT NON_TER 346 346
SQ SEQUENCE 346 AA; 39127 MW; A804B96BDDFA010 CRC64;

Query Match 12.5%; Score 112.5; DB 5; Length 346;
Best Local Similarity 23.7%; Pred. No. 2.4;
Matches 42; Conservative 33; Mismatches 67; Indels 35; Gaps 7;

QY 10 NTTVKEPILN-KDTGEVSELRKRVTVITQNGKMSSTIVSEEDPFIIPYKGLKGYOF 68
DB 156 STKTEYAKAKNAVYKAKNAVQKNAVLAKEASS-----YNYIL----- 197
QY 69 DQWEISGF--EGKQAG-----YVNLSDPTIKVPFKIEKKEENKPTPDVSKKON 121
DB 198 -GWERGGGVPRHKKEENMLSHLYVSSKQENISKENDVLDL-KEEBAETGQBLEEK 255
QY 122 PQVNHSQLNESHKREDLQREHESQKSDSTQVATV-----LDKNISKSTNN 171
DB 256 EETSESEINDEBESEBESEBESEBENDNKGQAKGQSDQKEDVKAQNLISKQNNN 312

RESULT 9
ID Q25995 PRELIMINARY: PRT: 354 AA.
AC Q25995:
DT 01-NOV-1996 (TRENBLREL. 01, Created)
DT 01-NOV-1996 (TRENBLREL. 01, Last sequence update)
DT 01-OCT-2002 (TRENBLREL. 22, Last annotation update)
DB Antigen.
OS Plasmodium falciparum (isolate NF54).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5843;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NF54;
RX MEDLINE=95198774; PubMed=7891748;
RA McColl D.J., Silva A., Foley M., Kun J.F., Favaloro J.M.,
RA Thompson J.K., Marshall V.M., Coppel R.L., Kemp D.J., Anders R.F.;
RT "Molecular variation in a novel polymorphic antigen associated with
RT Plasmodium falciparum merozoites."
RL Mol. Biochem. Parasitol. 68:53-67(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=NF54;
RX MEDLINE=98156743; PubMed=9497029;
RA McColl D.J., Anders R.F.;
RT "Conservation of structural motifs and antigenic diversity in the
RT Plasmodium falciparum merozoite surface protein-3 (MSP-3).";
RL Mol. Biochem. Parasitol. 90:21-31(1997).
DR EMBL; L28825; AAC09377.1;
SQ SEQUENCE 354 AA; 40119 MW; 3A7256152F48B527 CRC64;

Query Match 12.5%; Score 112.5; DB 5; Length 354;
Best Local Similarity 22.4%; Pred. No. 2.5;
Matches 41; Conservative 35; Mismatches 62; Indels 45; Gaps 7;

QY 10 NTTVKEPILN-KDTGEVSELRKRVTVITQNGKMSSTIVSEEDPFIIPYKGLKGYOF 68
DB 153 STKTEYAKAKNAVYKAKNAVQKNAVLAKEASS-----YNYIL----- 194
QY 69 DQWEISGF--EGKQAG-----YVNLSDPTIKVPFKIEKKEE-----E 108
DB 195 -GWERGGGVPRHKKEENMLSHLYVSSKQENISKENDVLDLKEEBAETGQBLEEK 253
QY 109 NKPTPDVSKKONPQVNHSQLNESHKREDLQREHESQKSDSTQVATVLDKNISKST 168
DB 254 EETSESEINDEBESEBESEBESEBENDNKGQAKGQSDQKEDVKAQNLISKQ 308
QY 169 TNN 171

```







RL Mol. Biochem. Parasitol. 68:53-67(1994).  
 [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=FC27;  
 RA MEDLINE=98156743; PubMed=9497029;  
 RX McColl D.J., Anders R.F.;  
 RT "Conservation of structural motifs and antigenic diversity in the Plasmodium falciparum merozoite surface protein-3 (MSP-3)."  
 RL Mol. Biochem. Parasitol. 90:21-31(1997).  
 DR EMBL; L07944; AAC09378.1; -  
 KW Signal.  
 FT SIGNAL 1 25 POTENTIAL.  
 FT CHAIN 26 380 POLYMORPHIC ANTIGEN.  
 SQ SEQUENCE 380 AA; 43290 MW; 0986CA1393094CA2 CRC64;  
 Query Match 12.3%; Score 110; DB 5; Length 380;  
 Best Local Similarity 24.2%; Pred. No. 4;  
 Matches 45; Conservative 34; Mismatches 59; Indels 48; Gaps 9;  
 DB 179 FSRKT--KEYAEKAKNAVEKAKNAVQKANKAVLKAKKAS-----YDYIL----- 221  
 QY 66 YQPGWEISGF--RGKQAG-----YVNLISKQTFIKVFKKIEKKKEENKPTFVSXK 118  
 DB 222 ---GWEFGGVPFRHKKENMLSHLYSSKDKENISKENDVLDK-KEEAEETEERLE 276  
 QY 119 KDNFVHNSQAN-----ESHKEDLQREHSQSKDSTQVATATVLDKNNISS 165  
 DB 277 EKNEETETSEISDEEBEERBEERKEENKKEQEKQENNDQKDYBA-----QNLIS 331  
 QY 166 KSTNN 171  
 DB 332 KNQNN 337  
 RESULT 17  
 ID 081436 PRELIMINARY; PRT; 3008 AA.  
 AC 081436;  
 DT 01-MAR-2003 (TREMBlrel. 23, Created)  
 DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)  
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)  
 DE Hypothetical protein.  
 GN PF80325W.  
 OS Plasmodium falciparum (isolate 3D7).  
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
 NCBI\_TaxID=36329;  
 RX SEQUENCE FROM N.A.  
 RA Devlin K., Baker S., Davies P., Mungai K., Berriman M., Pain A., Hall N., Bowman S., Churcher C., Quail M., Barrett B.;  
 RL Submitted (SEP-2002) to the EMBL/Genbank/DBJ databases.  
 [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22255708; PubMed=12368867;  
 RA Hall N., Pain A., Berriman M., Churcher C., Harris B., Harris D., Mungall K., Bowman S., Atkin R., Baker S., Barron A., Brooks K., Buckee C.O., Burrows C., Cherevach I., Chillingworth C., Chillingworth T., Christodoulou Z., Clark L., Clark R., Corton C., Cronin A., Davies R., Davis P., Dear P., Dearden F., Doggett J., Feltwell T., Goble A., Goodhead I., Gwilliam R., Hamlin N., Hance Z., Harper D., Hauser H., Hornsby T., Holroyd S., Horrocks P., Humphray S., Jagsels K., James K.D., Johnson D., Kerhornou A., Knights A., Kontorov B., Kyes S., Larke N., Lawson D., Lemard N., Line A., Maddison M., McLean J., Mooney P., Moule S., Murphy L., Oliver K., Omond D., Price C., Quail M.A., Rabinowitsch E., Rajandream M.A., Rutter S., Rutherford K.M., Sanders M., Simmonds M., Seeger K., Sharp S., Smith R., Squares R., Squares S., Stevens K., Taylor K., Tivey A., Unwin L., Whitehead S., Woodward J., Yulston J.E., Craig A., Newbold C., Barrett B.G.;  
 RL "Sequence of Plasmodium falciparum chromosomes 1, 3-9 and 13.";  
 RT Nature 419:527-531(2002).

DR EMBL; AL929351; CAD51431.1; -  
 KW Hypothetical protein.  
 SQ SEQUENCE 3008 AA; 356023 MW; 60BCBEE15C599B4 CRC64;  
 Query Match 12.2%; Score 109.5; DB 5; Length 3008;  
 Best Local Similarity 24.6%; Pred. No. 39;  
 Matches 52; Conservative 37; Mismatches 71; Indels 51; Gaps 10;  
 QY 1 KIVVDFPANTTVKFFILNKDTGVESEIK-----PHEVTVTIQNGKMSSTIVSE 50  
 DB 2218 KVLIEKCESLSTIKESHED-----EIKDDNNNNNNNNKHGYNNIDHKNINKEVTT 2272  
 QY 51 BDF-----ILPVYKGLBKGYQPDGWEISGFEKGKQAGYVNLISK---DTFIK 96  
 DB 2273 QDANSNPNLNNITIQKKGRNTRNSKVIFCEEIQVRE-----YDIEISKIEKFGASIGP 2326  
 QY 97 VPKIIEKKEEENKPTFVSXKKNPQVNHSG--LNSHREKEDLQREH-----SQKSDS 149  
 DB 2327 VFTD--EENKKEENKN--EVNKKKEENKKEENKKEENKKEENKKEENKKEEN 2383  
 QY 150 TKDVTATVLDKNNISSK-----STNNPNK 174  
 DB 2384 KKEVKNKEENKKEENKKEENKKEENKKEENK 2414  
 RESULT 18  
 ID 096229 PRELIMINARY; PRT; 951 AA.  
 AC 096229;  
 DT 01-MAY-1999 (TREMBlrel. 10, Created)  
 DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)  
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)  
 DE Hypothetical protein.  
 GN PF80680W.  
 OS Plasmodium falciparum (isolate 3D7).  
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
 NCBI\_TaxID=36329;  
 RX SEQUENCE FROM N.A.  
 RA Gartner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W., Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K., Riesen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S., Pater M.-S., Nene V., Shelton S.J., Sun B., Peterson J., Angiolini S., Chara M., Allen J., Selengut J., Hatt D., Mather M.W., Valdivya A.B., Martin D.M.A., Fairclamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A., McFadden G.I., Cummings L.M., Subramanian G.M., Mungall C., Venter J.C., Carucci D.J., Hoffman S.L., White O., Davies R.W., Praser C.M., Adams M.D., Venter J.C., Hoffman S.L.;  
 RL "Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.";  
 RT Science 282:1126-1132(1998).  
 [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=3D7;  
 RX MEDLINE=22255705; PubMed=12368864;  
 RA Gartner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W., Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K., Riesen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S., Pater M.-S., Nene V., Shelton S.J., Sun B., Peterson J., Angiolini S., Chara M., Allen J., Selengut J., Hatt D., Mather M.W., Valdivya A.B., Martin D.M.A., Fairclamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A., McFadden G.I., Cummings L.M., Subramanian G.M., Mungall C., Venter J.C., Carucci D.J., Hoffman S.L., White O., Davies R.W., Praser C.M., Adams M.D., Venter J.C., Hoffman S.L.;  
 RL "Genome sequence of the human malaria parasite Plasmodium falciparum.";  
 RT Nature 419:498-511(2002).  
 DR EMBL; AE001410; AAC71925.2; -  
 KW Hypothetical protein.  
 SQ SEQUENCE 951 AA; 112486 MW; AC8DB89358A84F4F CRC64;  
 Query Match 12.2%; Score 109; DB 5; Length 951;  
 Best Local Similarity 26.1%; Pred. No. 12;  
 Matches 49; Conservative 32; Mismatches 49; Indels 58; Gaps 11;

QY 5 KDPARTTVEKFIPLKDTGSEVSELPKPRVT-VTIONGKEMSGTIVSEDFILPVTKGEIE 63  
 DB 115 KODNNNNNGTKQIEBKNNKIKNSDL--HRONELALQSGK-----NEODI-----NNVE 159  
 QY 64 KGYQPDGHEI--SGREGKDDACGYVNLKDTFIKVFVKIEEKE-----EENKP 111  
 DB 160 KGRQ---DISNSNENKDD-----VKGVLEKEKKEKKEKISDHYKEENK 203  
 QY 112 TFD-----VSKKDNPNQVNSQLNESHREKDLOR--EHSOKSOSTVDVATVLDKNNISK 166  
 DB 204 SDDHKVNEKNSDDHKVNEKNSDDHKIEBKVKEVBEHEDEE-----DKKKEKSE 254  
 QY 167 STTNPNK 174  
 DB 255 NKNKDENK 262

## RESULT 19

QY 077355 PRELIMINARY; PRT; 600 AA.  
 AC 077355;  
 DT 01-NOV-1998 (TrEMBLrel. 08, Created)  
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)  
 DE Hypothetical 71.7 kDa protein.  
 GN PFC0465C, MAJ3P4.20.  
 OS Plasmodium falciparum (isolate 3D7).  
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
 CX NCBI\_TaxId=36329;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=3D7;  
 RX MEDLINE=99376085; PubMed=10448855;  
 RA Bowman S., Lawson D., Basham D., Brown D., Chillingworth T., Churcher C.M., Craig A., Davies R.M., Devlin K., Fellwell T., Gentles S., Gilliam R., Hamlin N., Harris D., Holtroyd S., Hornaby T., Horrocks P., Jagels K., Jassal B., Kyes S., McLean J., Moulé S., Mungall K., Murphy L., Oliver K., Quail M.A., Rajandream M.-A., Rutter S., Skelton J., Squares R., Squares S., Sulston J.E., Whitehead S., Woodhead J.R., Newbold C., Barrett B.G.;  
 RA "The complete nucleotide sequence of chromosome 3 of Plasmodium falciparum."  
 RT Nature 400:532-538(1999).  
 RL EMBL: AL008970; CA15610.2;  
 DR InterPro: IPR002483; PFI.  
 DR Pfam: PF01480; PFI; 1.  
 DR SMART: SM00311; PFI; 1.  
 KW Hypothetical protein.  
 SQ SEQUENCE 600 AA; 71663 MW; 57EAB42565CAD64C CRC64;

Query Match 12.1%; Score 108.5; DB 5; Length 600;  
 Best Local Similarity 29.3%; Pred. No. 8.3;  
 Matches 54; Conservative 22; Mismatches 47; Indels 61; Gaps 12;

QY 47 IVSEEDFLPVY-----KGELEKQVPGMEISGREGK-----DAGVYVNLKDPFIKRV 97  
 DB 60 ILGFEDDLIYECISQLKQSKERK--DGEEDKYINAKKLKINLGFIGNKKSDFIEEL 116  
 QY 98 FKKI--BEKKE-----ENKPTFVSK-KKDNPNQVNSQLN-----SHR 135  
 DB 117 LELLINKEKEHIDTLANK--TNDIKVKNENENINENYNNENKDKISNKEKHYSHON 175  
 QY 136 E-----DLQREH-----SOKSDSTK---DVATVLDKNNISKSTTN 170  
 DB 176 EHNINNVMLKKEKEYTDIGDRKGRKHSQSQSDSYKCKPFRKRTSIR-SLSNRYDE 234  
 QY 171 NPNK 174  
 DB 235 KTNK 238

## RESULT 20

ID 09FUK9 PRELIMINARY; PRT; 470 AA.  
 AC 09FUK9;  
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)  
 DE Gb|AAP20218.1.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC Eustida II; Brassicales; Brassicaceae; Arabidopsie.  
 CX NCBI\_TaxId=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Columbia;  
 RX MEDLINE=99087489; PubMed=9872454;  
 RA Nakamura Y., Sato S., Asamizu E., Kaneko T., Kotani H., Miyajima N., Tabata S.;  
 RA "Structural analysis of Arabidopsis thaliana chromosome 5. VII. Sequence features of the regions of 1,013,767 bp covered by sixteen physically assigned P1 and TAC clones."  
 RL DNA Res. 5:297-308(1998).  
 RX EMBL: AB015468; BAB10694.1;  
 SQ SEQUENCE 470 AA; 53758 MW; 6D686CE72B35AC54 CRC64;  
 Query Match 12.0%; Score 107.5; DB 10; Length 470;  
 Best Local Similarity 20.1%; Pred. No. 7.5;  
 Matches 36; Conservative 32; Mismatches 66; Indels 45; Gaps 5;  
 QY 29 KPRVTVTIONGKEMSGTIVSEDFILPVYKGELEKGYQPDGMEISGFE-----GKK 80  
 DB 82 RENVITDYNNSNGSK-----YGDILARRIYDE-EATGSQSAQRIDHPNOK 129  
 QY 81 DAGVYVNLKDPFIKVFVKIEEKEENKPTFVSKKKN----- 121  
 DB 130 NVGITEKAPENSPIBETSHRVDNKRINNOKFTAKSSENAVRSFGADHRAEVMGR 189  
 QY 122 PNVNSQLN-----SHRKELOREBSQKSDSTVDVATVLDKNNISKSTTNPNK 174  
 DB 190 PMENDDQVQTESAKSHKRENVYTSSEKPRDQGVKTEAKDKRKEKKEKTESINK 248  
 RESULT 21  
 ID 09U4U6 PRELIMINARY; PRT; 948 AA.  
 AC 09U4U6;  
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)  
 DE Ornithine decarboxylase.  
 GN ODC.  
 OS Plasmodium falciparum.  
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
 CX NCBI\_TaxId=5833;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=FCP1/RSa;  
 RA Birholtz U., Joubert F., Neitz A.W.H., Louw A.L.;  
 RA "Molecular characterisation of Plasmodium falciparum ornithine decarboxylase cDNA obtained by RACE."  
 RT Submitted (Apr-1999) to the EMBL/Genbank/DBJ databases.  
 RX EMBL: AF139900; AAF14518.1;  
 DR HSPF: P07805; I137.  
 DR InterPro: IPR000183; Decarboxylase2.  
 DR Pfam: PF02784; Orn\_Arg\_dec\_N; 1.  
 DR Pfam: PF02784; Orn\_DAP\_Arg\_dec; 1.  
 DR PRINTS: PR01179; ODACRBLASE.  
 SQ SEQUENCE 948 AA; 110350 MW; 43F103DB83F12835 CRC64;  
 Query Match 12.0%; Score 107.5; DB 5; Length 948;  
 Best Local Similarity 22.0%; Pred. No. 16;  
 Matches 54; Conservative 32; Mismatches 71; Indels 89; Gaps 11;  
 QY 4 VKDPARTTVEKFIPLKDTGSEVSEL-----KPRVTVTIONGKEMSGTIVSEDFI 54



Best Local Similarity 25.4%; Pred. No. 50;  
Matches 48; Conservative 31; Mismatches 71; Indels 39; Gaps 8;

QY 2 IYVQPARNTTKEFLANDTGVSELSKPHRYVTYIQNGKMSSTVSEDFILPYKGB 61  
||| : : : : :  
DB 1639 IYVNRGKSDSIKE-----GSEDEK-----TVEINGSHRLSTEGSKD-----CK 1678

QY 62 LKQGFQDMEWISGFEKGD-----AGVIVLSKDTPIKPYFKKIEKKEENRP 111  
||| : : : : :  
DB 1679 IEEGR-EKGENSTYKESKDDEKIEGMEKENSTKSSKDGKINELHGDKEATMEGSKD 1736

QY 112 --TFPVSK-KKONPOVNSQLMBSHRKDLQ-----REBSQKSDTKDYATVLDKNNIS 164  
||| : : : : :  
DB 1737 GGTNSTGDSKDSKSVKVEINGVDKSLKDDSKGKDINEINNGKSDVKDNTBIQGDNSL 1796

QY 165 SKSTTNPN 173  
||| : : : : :  
DB 1797 TMTSTSEPN 1805

RESULT 25  
Q8IDA0 PRELIMINARY; PRT; 3127 AA.  
AC Q8IDA0;  
DT 01-MAR-2003 (TREMBLrel. 23, Created)  
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)  
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)  
DE Guanylyl cyclase (EC 4.6.1.2).  
GN GC-BETA.  
OS Plasmodium falciparum (isolate 3D7).  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.  
RN NCBI\_TaxID=36329;  
RM [1]  
RP SEQUENCE FROM N.A.  
RA Harris B., Lennard N., Clark L., Line A., Barron A., Corton C.,  
RA Berriman M., Pain A., Hall N., Atkin R., Chillingworth C., Doggett J.,  
RA Omond D., Sanders M., Hayes R., Hall S., Quail M., Barrell B.,  
RA Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AL844509; CAD52725.1; -.  
KW Lyase.  
SQ SEQUENCE 3127 AA; 370597 MW; F0375C72B9007560 CRC64;

Query Match 11.7%; Score 105; DB 5; Length 3127;  
Best Local Similarity 23.7%; Pred. No. 84;  
Matches 55; Conservative 36; Mismatches 73; Indels 68; Gaps 10;

QY 9 RNTTVK--EFLINKDTGKSELK-----HRYVTYIQNG-----KMSSTVSE 50  
||| : : : : :  
DB 1950 RNTTKYKSDYLLIDMKGESKFKRNTSYVLBSPLHLIDIVNNIKRKKKEIKTIYSD 2009

QY 51 BDFILPYKGBLEKGYQPDGWSISG---FEGKGDAGYVNLKSDTFIKVFPKIEKKEE 107  
||| : : : : :  
DB 2010 DMFSPVNIKEYNVNEQERKEIVGNLSYDTKTKIIPPFIKFKEGRIK--KKIEKKEK 2067

QY 108 ENK-----PTF-----DVSKKDKNPQVNSQL----- 129  
||| : : : : :  
DB 2068 EKGNNNNFLYNDYSSYSSPYGNGENNENVIKIRERKDFQKKDPHPNFSKPLANNY 2127

QY 130 ---NESHRK---EDLQREHSQKSDSTKD-VTATVLDKNNISKSTNNPNK 174  
||| : : : : :  
DB 2128 PMKNNKNNKNNKNNKNNRNEYPNYTTSSKDGVSYNFLSDLSFSSDMEYSDBN 2179

RESULT 26  
Q94C59 PRELIMINARY; PRT; 540 AA.  
AC Q94C59;  
DT 01-DEC-2001 (TREMBLrel. 19, Created)  
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)  
DE Putative phosphatidyl-inositol-transfer protein.  
GN T518.14.  
RA Arabidopsis thaliana (Mouse-ear cress).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC Eumetazoa; Brasiaceae; Brasiaceae; Arabidopsis.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Yamada K., Liu S.X., Sakano H., Pham P.K., Banh J., Chung M.K.,  
RA Goldsmith A.D., Lee J.M., Quach H.L., Tang C., Toriumi M., Yu G.,  
RA Bower L., Carrinci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J.,  
RA Jones T., Kamiya A., Karlun-Neumann G., Kawai J., Kim C., Koesema B.,  
RA Lam B., Lin J., Meyers M.C., Miranda M., Narusaka M., Nguyen M.,  
RA Palm C.J., Sakurai T., Satou M., Seki M., Shim P., Southwick A.,  
RA Shinozaki K., Davis R.W., Becker J.R., Theologis A.,  
RT "Full Length cDNA of gene T518.14 (GI:4587525)." ;  
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AY035162; AAK59666.1; -.  
DR InterPro: IPR001251; CRAL TRIO.  
DR InterPro: IPR001071; RetBind/tocTrans.  
DR Pfam: PF00650; CRAL TRIO; 1.  
DR Pfam: PF03765; CRAL TRIO; N; 1.  
DR PRINTS: PR00180; CRETINALDHP.  
DR SMART: SM00516; SRC14; 1.  
DR PROSITE: PS00191; CRAL TRIO; 1.  
SQ SEQUENCE 540 AA; 61166 MW; 0C2592EB18ACFAC6 CRC64;

Query Match 11.6%; Score 104; DB 10; Length 540;  
Best Local Similarity 25.3%; Pred. No. 15;  
Matches 40; Conservative 28; Mismatches 64; Indels 26; Gaps 6;

QY 11 TVYKEFLINKDTGSEVSELSKPHRYVTYIQNGKMSSTVSEDFILPYKGBLEKGYQPDG 70  
||| : : : : :  
DB 27 TVYKA-VSEETKVEDESKP-----EGVKSASFKESDPSFALUKESBK----- 70

QY 71 WEISGFEKGDAGYVNLKSDTFIKVFPK--IEBKKEBNKPTDVSCKKONPOVNS 127  
||| : : : : :  
DB 71 -ALSDKSLLEFAYVN-----TLTKTKKSSPMKKGKGRVYPAEVEKKE--EAAB 123

QY 128 QNESHRKEDLQREHSQKSDSTKDYATVLDKNNIS 165  
||| : : : : :  
DB 124 KVEEKSEAVVTEAPKATVEAVVTEETIIPREVT 161

RESULT 27  
Q9SA84 PRELIMINARY; PRT; 540 AA.  
AC Q9SA84;  
DT 01-MAY-2000 (TREMBLrel. 13, Created)  
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)  
DE T518.14 protein (Hypothetical protein).  
GN T518.14 OR ATG30690.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC Eumetazoa; Brasiaceae; Brasiaceae; Arabidopsis.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CV. Columbia;  
RA Vysotskaia V.S., Schwartz J.R., Yu G., Toriumi M., Lenz C., Liu S.,  
RA Li J., Kremenetskaia I., Lurov J., Ngan I., Gonzalez A., Aitafi H.,  
RA Araujo R., Chao Q., Conn L., Conway A.B., Dunn P., Hansen N.,  
RA Huitzer L., Kim C., Palm C., Rowley D., Shim P., Walker M.,  
RA Davis R.W., Becker J.R., Federspiel N.A., Theologis A.,  
RT "Arabidopsis thaliana chromosome 1 BAC T518 sequence." ;  
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Yamada K., Chan M.M., Chang C.H., Dale J.M., Hsuan V.W., Lee J.M.,  
RA Quach H.L., Tang C.C., Toriumi M., Wallender E.K., Wong C., Wu H.C.,  
RA Yu G., Yuan S., Chen H., Cheuk R., Jones T., Kim C.J., Nguyen M.,  
RA Palm C.J., Shim P., Southwick A., Timp M.G., Wu T., Davis R.W.,  
RA Becker J.R., Theologis A.,

RT "Arabidopsis Open Reading Frame (ORF) Clones."  
 RL Submitted (OCT-2002) to the EMBL/Genbank/DBJ databases.  
 DR EMBL: AC007060; AAD25756.1; -;  
 DR EMBL: BT000959; AAM41359.1; -;  
 DR InterPro: IPR001251; CRAL\_TRIO.  
 DR InterPro: IPR001071; RetBind/Loctrans.  
 DR Pfam: PF00650; CRAL\_TRIO; 1.  
 DR Pfam: PF03765; CRAL\_TRIO\_N; 1.  
 DR PRINTS: PR00180; CRETINALDHP.  
 DR SMART: SM00516; SEC14; 1.  
 DR PROSITE: PS0191; CRAL\_TRIO; 1.  
 KW Hypothetical protein.  
 SQ SEQUENCE 540 AA; 61189 MW; 0C2590D518ACFB58 CRC64;

Query Match 11.6%; Score 104; DB 10; Length 540;  
 Best Local Similarity 25.3%; Pred. No. 15;  
 Matches 40; Conservative 28; Mismatches 64; Indels 26; Gaps 6;

QY 11 TTVEFLINKTGEVSELKPHRYVTITQNGKEMSTIYSEDFILPVYKGLKGYQFDG 70  
 DB 27 TTVA-VVEFKVDESKP-----EGVEKSAFKEESDFADLKESEK----- 70  
 QY 71 WEISGFEKQAGVYINLTKDTPIKVPFK--IREKKEENKPTFDVSKKKNQVNH 127  
 DB 71 -ALSDLSKLEBAIVD---TLTKKSSSPMKKEEVVKEPAVEKKE--EAAAE 123  
 QY 128 QLNESHKEDLQREHSOKSDSTKDTATVLDKNISS 165  
 DB 124 KVEBKSEAVTTEAPAKATVEAVTTEIIPKEVTT 161

## RESULT 28

044016 PRELIMINARY; PRT; 325 AA.

AC 044016;  
 DT 01-JUN-1998 (TREMBLrel. 06, Created)  
 DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)  
 DT 01-OCT-2002 (TREMBLrel. 22, Last annotation update)  
 DE G5 ORF.  
 OS Dictyostelium discoideum (Slime mold).  
 OC Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.  
 NC NCB1\_TaxID=44689;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=NC4;  
 RX MEDLINE=98198836; PubMed=9539429;  
 RA Rieben W.K. Jr., Gonzales C.M., Gonzalez S.T., Pilkington K.J.,  
 RA Kiyosawa H., Hughes J.B., Welker D.L.;  
 RT "Dictyostelium discoideum nuclear plasmid dpgs is a chimera related to  
 the Ddpl and Ddpg2 plasmid families";  
 RL Genetics 148:1117-1125(1998).  
 DR EMBL: U00796; AACT18634.1; -;  
 SQ SEQUENCE 325 AA; 38448 MW; 69A43D0C632058A6 CRC64;

Query Match 11.5%; Score 103.5; DB 5; Length 325;  
 Best Local Similarity 23.8%; Pred. No. 9.5;  
 Matches 43; Conservative 25; Mismatches 50; Indels 63; Gaps 9;

QY 31 HRVVTITQNGKEMSTIYSEDFILPVYK-GEL--EKGYQFDGKISGFEK----- 79  
 DB 57 HRTTSTIN--RFSVKKIGDESKLFRISKNGELIVNLLEFDNPHIK--EGKHLKSKMF 112  
 QY 80 ---KDAGV-----INLSDTPIKPV-----FK----- 100  
 DB 113 NHIDSGYATNEIEIFLBSCTLCKETTAQTKRSYKKNINLTPREBEHEBEHEBE 172  
 QY 101 --IEKKEENKPTFDVSKKKNQVNHSHRKE-----DLQREHSOKSDSTK 151  
 DB 173 EEEBQEEVEVKPTISEEBBETPAVSEBKKEBEHEBEHEBEHEBEHEBEHEBE 232  
 QY 152 D 152  
 DB 233 D 233

## RESULT 29

081226 PRELIMINARY; PRT; 238 AA.

AC 081226;  
 DT 01-MAR-2003 (TREMBLrel. 23, Created)  
 DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)  
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)  
 DE Hypothetical protein.  
 GN PF10765W.  
 OS Plasmodium falciparum (isolate 3D7).  
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
 NC NCB1\_TaxID=36329;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22255708; PubMed=12368867;  
 RA Hall N., Pain A., Berriaman M., Churcher C., Harris B., Harris D.,  
 RA Mungall K., Bowman S., Atkin R., Baker S., Barron A., Brooks K.,  
 RA Buckee C.O., Burrows C., Cherevach I., Chillingworth C.,  
 RA Chillingworth T., Christodoulou Z., Clark L., Clark R., Corton C.,  
 RA Cronin A., Davies R., Davis P., Dear P., Dearden F., Doggett J.,  
 RA Felwell T., Goble A., Goodhead I., Gilliam R., Hamlin N., Hance Z.,  
 RA Harper D., Hauser H., Hornsby T., Holroyd S., Horrocks P.,  
 RA Humphrey S., Jagers K., James K.D., Johnson D., Kethornou A.,  
 RA Knights A., Kontorov B., Kyes S., Larke N., Lawson D., Leonard N.,  
 RA Line A., Maddison M., McLean J., Mooney P., Moule S., Murphy L.,  
 RA Oliver K., Ormond D., Price C., Quail M.A., Rabinovitch E.,  
 RA Rajandream M.A., Rutter S., Rutherford K.M., Sanders M., Simmonds M.,  
 RA Seeger K., Sharp S., Smith R., Squares R., Squares S., Stevens K.,  
 RA Taylor K., Tivey A., Unwin L., Whitehead S., Woodward J.,  
 RA Sulston J.B., Craig A., Newbold C., Barrell B.G.;  
 RT "Sequence of Plasmodium falciparum chromosomes 1, 3-9 and 13.";  
 RL Nature 419:527-531(2002).  
 DR EMBL: AL29356; CAD51839.1; -;  
 KW Hypothetical protein.  
 SQ SEQUENCE 238 AA; 28481 MW; DA175A2B5A109B49 CRC64;

Query Match 11.5%; Score 103; DB 5; Length 238;  
 Best Local Similarity 26.8%; Pred. No. 7.4;  
 Matches 52; Conservative 30; Mismatches 74; Indels 38; Gaps 11;

QY 1 KIVVDFPANTTVKFIINKDGEVSELKPHRYVTITQNGKEMSTIYSEDFILPVYK 60  
 DB 22 KQIRKEKTTTSQNL--KEKEKQETKQHD-----ENKSNVNVNSSEFP--KLIYKN 72  
 QY 61 ELEKGYQFDGWEISG--FEKQDA-----GYVINLSDTPIKVPFKIEKKEENK 110  
 DB 73 ILKQVYHBEKPKQKQENIEESKKATSRKIKGLNPKLKNLIEKVIVK--KAPLEAK 129  
 QY 111 PTFVSKKKN--PQV--NH--SQLNESHKED-----LQREHSOKSDSTKDTATVLD 159  
 DB 130 NTVFSLKRYVTVPVHISNNHNSQNDHNKKQDKNIKLLKLEKSNKQDN-----NINN 184  
 QY 160 KNKISSKSTTNPN 173  
 DB 185 KKNNNQKCDYTNLN 198  
 RESULT 30  
 093424 PRELIMINARY; PRT; 385 AA.  
 ID 093424;  
 AC 093424; P90801;  
 DT 01-JAN-1999 (TREMBLrel. 09, Created)  
 DT 01-JAN-1999 (TREMBLrel. 09, Last sequence update)  
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)  
 DE Hypothetical glycine-rich 37.0 kDa protein f02a10.2 in chromosome V  
 DE Precursor.  
 GN E02A10.2.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;  
 OC Rhabditidae; Pelodetidae; Caenorhabditis.  
 NC NCB1\_TaxID=6239;

RM [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RA Thomas K.;  
RL Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.  
RN (2)  
RP REVISIONS.  
RC STRAIN-BRISTOL N2;  
RA Jones S.J.M.;  
RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.  
DR EMBL; 281053; CAB02877.1; -.  
DR HSSP; P10968; 2CWG.  
DR WormPep; E02A10.2; CE09116.  
DR InterPro; IPR002962; Eggshell.  
DR PRINTS; PR01228; EGGSHELL.  
KW Hypothetical protein; Signal; Transmembrane.  
FT SIGNAL 1 17  
FT CHAIN 18 385  
FT TRANSMEM 165 185  
FT DOMAIN 23 190  
FT SEQUENCE 385 AA; 36963 MW; 32AB25AE9B5073FE CRC64;  
SQ

Query Match 11.4%; Score 102.5; DB 5; Length 385;  
Best Local Similarity 26.2%; Pred. No. 13;  
Matches 48; Conservative 31; Mismatches 55; Indels 49; Gaps 9;

QY 1 KIVVDKPARNT-----TWKEFLIND-----TGEVSELPKPRVTVTQ- 38  
DB 215 RIILNENTKETALESKITKEKAGODYIVVCKEKAPPAFTATDDCSLOKENVHCTILR 274  
QY 39 -NGKMSSTIVSEEDFILPVYKGELEKGYQFDGMEISGFEKKDAGVNLNLSKDTFIKCV 97  
DB 275 INHKEVAEK--NEBDKKEEPKKEEKEVEKEK--EDBKDKD-----EP- 316  
QY 98 FKXIEKKEEKEKPTFDVSKKONPVNHSQLESKRKDLQREHSQKSDSTQVATV 157  
DB 317 -KKEEKEKEEOKK--EVEKKEE-----EKXDEPKKEEKKKEEKEDEVEKSEKCY 368  
QY 158 LDK 160  
DB 369 BEK 371

## RESULT 31

Q8IHW3 PRELIMINARY; PRT; 449 AA.  
AC Q8IHW3;  
DT 01-MAR-2003 (TRENBLREL. 23, Created)  
DT 01-MAR-2003 (TRENBLREL. 23, Last sequence update)  
DT 01-MAR-2003 (TRENBLREL. 23, Last annotation update)  
DE Hypothetical protein.  
GN PF11\_0413.  
OS Plasmodium falciparum (isolate 3D7).  
OC Bacteria; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
OX NCBI\_TaxID=36329;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=3D7;  
RA MEDLINE=22255705; PubMed=12368864;  
RA Gardner M.J., Hall N., Fung B., White O., Berriman M., Hyman R.W.,  
RA Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,  
RA Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,  
RA Chan M.-S., Nene V., Shallow S.J., Sub B., Peterson J., Angiuoli S.,  
RA Pereira M., Allen J., Selengut J., Haft D., Mather M.W., Valdivia A.B.,  
RA Martin D.M.A., Fairclimb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,  
RA McFadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,  
RA Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,  
RA Graer C.M., Barrall B.;  
RT "Genome sequence of the human malaria parasite Plasmodium  
falciparum.";  
RL Nature 419:498-511 (2002).  
DR EMBL; AB014842; AAN3596.1; -.

KW Hypothetical protein.  
SQ SEQUENCE 449 AA; 52812 MW; 85B62272D6257C68 CRC64;  
SQ

Query Match 11.4%; Score 102.5; DB 5; Length 449;  
Best Local Similarity 24.9%; Pred. No. 16;  
Matches 52; Conservative 30; Mismatches 58; Indels 69; Gaps 9;

QY 1 KIVVDKPARNTTKE--FLNKDGTGEVSELKPRVTVTQNGKMSSTIVSEEDFILPVY 58  
DB 31 KSVRDVADPLFVNBESYFENKEK-----KELKTLIKER--ILILY 70  
QY 59 -----KGELEKGYQFDGMEISGFEKKDAGVNLNLSKDTFIKPFKIE----- 102  
DB 71 NEQNEBEKKKKKKKQSD--ISKEKESKADLOKSESNNMLKKRKXIISESEETNINS 127  
QY 103 --EKKEE-----NKPTDV-----SKKONPVNHSQLESKRKDLQREH 143  
DB 128 DDEEHEEYQRRKKQKKNNSVSTLSLEKKKKRRDSESSNNNDYBDEDDDEQEE 187  
QY 144 SOKSDSTQVATVLDKNNISKSSTTNP 172  
DB 188 EEEES-----LSKKSKKKNSTTSP 207

## RESULT 32

Q98QAI PRELIMINARY; PRT; 622 AA.  
AC Q98QAI;  
DT 01-OCT-2001 (TRENBLREL. 18, Created)  
DT 01-OCT-2001 (TRENBLREL. 18, Last sequence update)  
DT 01-OCT-2002 (TRENBLREL. 22, Last annotation update)  
DE LIPOPROTEIN.  
GN MYP 4650.  
OS Mycoplasma pulmonis.  
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.  
OX NCBI\_TaxID=2107;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=UAB CTIP;  
RX MEDLINE=21267165; PubMed=11353084;  
RA Chambaud I., Hellig R., Ferris S., Barbe V., Samson D., Galisson F.,  
RA Moszer I., Dydvig K., Wroblewski H., Viari A., Rocha E.P.C.,  
RA Blanchard A.;  
RT "The complete genome sequence of the canine respiratory pathogen  
Mycoplasma pulmonis.";  
RL Nucleic Acids Res. 29:2145-2153 (2001).  
DR EMBL; AL445564; CAC13638.1; -.  
KW Mypulist; MYP 4650; -.  
SQ SEQUENCE 622 AA; 73762 MW; 42BD8930861960D CRC64;  
SQ

Query Match 11.4%; Score 102; DB 16; Length 622;  
Best Local Similarity 24.9%; Pred. No. 24;  
Matches 48; Conservative 35; Mismatches 66; Indels 44; Gaps 11;

QY 9 RNTYKEPIL-NKDTGEVSEL-----KPHRY--TWITONG--KEM--STI--V 48  
DB 26 KNDGAKNFVLSKSDLSISEKQFKNLIDKQKIIEGVITFINSETKIVYKKEITLNL 85  
QY 49 SEEDFILPVYKGELEKGYQFDGMEISGFEKKDAGVNLNLSKDTFIKPFKIEKKEE 108  
DB 86 SEDNIIIFSLINENNAKPOLDEF-----VSADKFKIKQFQINFGQTBQ 129  
QY 109 NKPTDVSKKCKD---NPVNHSQLNES--HRKEDLOREHSQKSDSTQVATVLDKN- 161  
DB 130 -KITDNISKEDEKKNKPNKDNNSNNSSDQKNDLQKNSDKLNDVQDEKANKENSNS 188  
QY 162 NISSKSTTNNPK 174  
DB 189 NDSKEKNDENTNK 201

## RESULT 33



Q9UBD4 PRELIMINARY; PRT; 1419 AA.  
AC Q9UBD4;  
DT 01-MAY-2000 (TREMBlrel. 13, Created)  
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)  
DE S-adenosylmethionine decarboxylase-ornithine decarboxylase.  
OS Plasmodium falciparum.  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
NCBI\_TaxID=5633;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=20179918; PubMed=10713131;  
RA Mueller S., Da'dera A., Luersen K., Wrenger C., Das Gupta R.,  
RA Madhubala R., Walter R.D.;  
RT "In the Human Malaria Parasite Plasmodium falciparum, Polyamines Are  
RT Synthesized by a Bifunctional Ornithine Decarboxylase, S-  
RT Adenosylmethionine Decarboxylase.";  
RL J. Biol. Chem. 275:8097-8102(2000).  
DR EMBL; AF094833; AAF00073.1; -.  
DR HSSP; P07805; 1P3T.  
DR InterPro; IPR000183; Decarboxylase2.  
DR InterPro; IPR001985; SAM decarbox.  
DR Pfam; PF002784; Orn\_Arg\_dec\_N; 1.  
DR Pfam; PF00278; Orn\_DAP\_Arg\_dec; 1.  
DR Pfam; PF01536; SAM decarbox; 1.  
DR PRINTS; PRO1179; ODAICRBLXASE.  
DR PRODOM; P002379; SAM decarbox; 1.  
SQ SEQUENCE 1419 AA; 166441 MW; 8842B97C0505686E CRC64;  
Query Match 11.4%; Score 102; DB 5; Length 1419;  
Best Local Similarity 22.0%; Pred. No. 58;  
Matches 52; Conservative 37; Mismatches 77; Indels 70; Gaps 11;  
QY 4 VKDPAKNTTVEKPIKNTQGEVSEI-----KPHVVTYTIQNGKMSSTIYSEDFI 54  
DB 1019 VFDSSNMGNFPIYIINLGGYPEBLEYDNAKKHDKIHCTSLDEIKKIDQKFLNEETFL 1078  
QY 55 LFPV-----KGELEKGYFGDGEISGFEKKD-----AGVYINLSKDTFIKVPFK 100  
DB 1079 KTKYGYVSFEKISLAIMNSIDHY-----FSHMKDNLRIVCERPIYVAASSTLAIVKIKGR 1134  
QY 101 -----IEKKKEENKPTF--DVSKKQDNQVNH-----SOLNESHKEDLQREHSOK- 146  
DB 1135 RPTQGIIMLKELKCHYDPLNPAQENKKQDETKINHNNDNDNDNDNDNDNDNNNNNNK 1194  
QY 147 -----SDST-----KDVATVLDK--NNIS-SKSTTNMNP 173  
DB 1195 GCGNIMDLITSTNDSTSKNDHSSSQVIONVCTITNDKGDNIKITHITINPNP 1250  
RESULT 34  
Q8IFNO PRELIMINARY; PRT; 1612 AA.  
AC Q8IFNO;  
DT 01-MAR-2003 (TREMBlrel. 23, Created)  
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)  
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)  
DE Hypothetical protein.  
OS Plasmodium falciparum (isolate 3D7).  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
NCBI\_TaxID=36329;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Hamlin N., Pain A., Berriman B., Hall N., Bowman S., Churcher C.,  
RA Harris B., Harris D., Lawson D., Quail M., Barrell B.;  
RT Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AL035477; CAD49270.1; -.  
KM Hypothetical protein.  
SQ SEQUENCE 1612 AA; 190946 MW; F00A408EB84521D07 CRC64;  
Query Match 11.4%; Score 102; DB 5; Length 1612;

Best Local Similarity 27.2%; Pred. No. 67;  
Matches 47; Conservative 29; Mismatches 73; Indels 24; Gaps 9;  
QY 9 RNTYKKEFLNKDTEGSELSKPHRYVTYTIQNG-KEMSSITVSE-----EDFILPVYKGL 62  
DB 1154 RNMVSNOVILNKTYDIDFCKENDKPVIKKIKIMTFVLELNYLDGFINSNINPFL 1213  
QY 63 EKGYPDGEISGFEKKDAGVYINLSKDTF-IKPEVFKIEKKKEENKPTDVSKKKN 121  
DB 1214 AKESYMYW-----KKQIN-VAKYNNKDDYDISFPKCIDENIKQN---YDINGQTNN 1262  
QY 122 PQVNHSQLNESHKEDLQREHSOKSDSTKDVATVLDKNISKSSTTNPNK 174  
DB 1263 NTVANNQTSN--HMENLHTDR-SKSNNNQDKTVYT---NEVSYLPIPIPKKK 1309  
RESULT 35  
Q9PPL5 PRELIMINARY; PRT; 312 AA.  
ID Q9PPL5  
AC Q9PPL5;  
DT 01-OCT-2000 (TREMBlrel. 15, Created)  
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)  
DE Putative membrane protein.  
CN C0692C.  
OS Campylobacter jejuni.  
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;  
OC Campylobacteriaceae; Campylobacter.  
NCBI\_TaxID=197;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX STRAIN=NCIC 11168; PubMed=10688204;  
RA MEDLINE=20150912; PubMed=10688204;  
RA Parthill J., Wren B.W., Mungall K., Ketley J.M., Churcher C.,  
RA Baeham D., Chillingworth T., Davies R.M., Fellwell T., Holroyd S.,  
RA Jagels K., Karlyshev A.V., Moute S., Pallen M.J., Penn C.W.,  
RA Quail M.A., Rajandream M.A., Rutherford K.M., van Vliet A.H.M.,  
RA Whitehead S., Barrell B.G.;  
RT "The genome sequence of the food-borne pathogen Campylobacter jejuni  
RT reveals hypervariable sequences.";  
RL Nature 403:665-668(2000).  
DR EMBL; AL139076; CAB72966.1; -.  
KM Complete proteome.  
SQ SEQUENCE 312 AA; 37221 MW; 0004FA7836A741B8 CRC64;  
Query Match 11.3%; Score 101.5; DB 16; Length 312;  
Best Local Similarity 25.0%; Pred. No. 13;  
Matches 46; Conservative 34; Mismatches 73; Indels 31; Gaps 9;  
QY 15 EFLNKDTEGSELSKPHRYVTYTIQNGKMSSTI---VSEEDFILPVYK-----GRL 63  
DB 5 DFIKERNQIRQKMLKFSR--AINQKPLDDDLRDEISSDILRRRFXKKTGNKFLERLD 61  
QY 64 KGYPDGEISGFEKKDAGVYINL---SKDTFIKVPFKIEKKKEENKPT---FDVSK 117  
DB 62 EYTSKTKTKSNYIYKED--LVNKLBEKQSLAKTITSKMERKKEBKTKTKQFLFSR 118  
QY 118 KQDN-----QVNHSQLNESHKEDLQREHSOKSDSTQDV--TATVLDKNNISSK--STT 169  
DB 119 KKANEIKNIQTKTQIQTNSQATQTTQKQKELNLSIKIQITERTKIQKPLIEKKLDVK 178  
QY 170 NNPN 173  
DB 179 NQPN 182  
RESULT 36  
Q9V7J0 PRELIMINARY; PRT; 382 AA.  
ID Q9V7J0  
AC Q9V7J0; Q9G081;  
DT 01-MAY-2000 (TREMBlrel. 13, Created)  
DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)  
DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)

DE CG6421 protein (Aspartyl beta-hydroxylase variant 2).  
 GN ASPH OR CG6421 OR CG18658.  
 RA Drosophila melanogaster (Fruit Fly).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephyroidea; Drosophilidae; Drosophila.  
 NCBI\_TaxID=7227;  
 RX STRAIN=Berkley;  
 MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amaratunga C., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blake R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abil J.P., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Bokoyev D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,  
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Houston K.A., Howland T.J., Wei M.-H., Ibegwam K.,  
 RA Jaitai M., Kalush F., Karpen G.H., Ke Z., Kennison K.A., Kechum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Laoko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Mckulov G., Misha N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Patel J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Part V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissenbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of Drosophila melanogaster.";  
 RL Science 287:2185-2195 (2000).  
 RN (2)  
 RP SEQUENCE FROM N.A.  
 RA Celniker S.E., Adams M.D., Krommiller B., Wan K.H., Holt R.A.,  
 RA Evans C.A., Gocayne J.D., Amaratunga C., Brandon R.C., Rogers Y.,  
 RA Barton J., An H., Baldwin D., Banzone J., Beeson K.Y., Busam D.A.,  
 RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,  
 RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresner D., Fartin D.,  
 RA Ferreira S., Friese E., Galie R.F., Garg N.S., George R.A.,  
 RA Gonzalez M., Houck J., Hoskins R.A., Houston D., Howland T.J.,  
 RA Ibegwam C., Jaitai M., Kruse D., Li P., Mattei B., Moshrefi A.,  
 RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,  
 RA Paclet J., Paragas V., Park S., Patel S., Pfeiffer B.,  
 RA Phouanavong S., Pittman G.S., Part V., Richards S., Scheeler F.,  
 RA Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,  
 RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;  
 RT "Sequencing of Drosophila melanogaster genome.";  
 RL Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.  
 RN (3)  
 RP SEQUENCE FROM N.A.  
 RA Miera S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,  
 RA Hildebeck P., Huang Y., Kaminker J.S., Prochnik S.B., Smith C.D.,  
 RA Tupy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.B.,  
 RA Clamp M., Drysdale R., Emmert D., Friese E., de Grey A., Harris N.,  
 RA Krommiller B., Marshall B., Milburn G., Richter J., Russo S.,

RA Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,  
 RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;  
 RT "Annotation of Drosophila melanogaster genome.";  
 RL Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.  
 RN (4)  
 RP SEQUENCE FROM N.A.  
 RA Adams M.D., Celniker S.B., Gibbs R.A., Rubin G.M., Venter J.C.;  
 RL Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.  
 RN (5)  
 RP SEQUENCE FROM N.A.  
 RA FlyBase;  
 RL Submitted (SEP-2002) to the EMBL/Genbank/DBJ databases.  
 RN (6)  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE=20564328; PubMed=10956665;  
 RA Dinchuk J.B., Henderson N.L., Burn T.C., Huber R., Ho S.P., Link J.,  
 RA O'Neil K.T., Focht R.J., Scully M.S., Hollis J.M., Hollis G.F.,  
 RA Friedman P.A.;  
 RT "Aspartyl beta-Hydroxylase (Asph) and an Evolutionarily Conserved  
 RT Isoform of Asph Missing the Catalytic Domain Share Exons with  
 RT Junction.";  
 RL J. Biol. Chem. 275:39543-39554 (2000).  
 DR EMBL: AB003808; AAC58063.2; -;  
 DR EMBL: AF289494; AAC40807.1; -;  
 DR FlyBase: FBgn0034075; Asph.  
 SQ SEQUENCE 382 AA; 43287 MW; 60E5C03ABDFC6E8B CRC64;  
 Query Match 11.3%; Score 101.5; DB 5; Length 382;  
 Best Local Similarity 24.5%; Pred. No. 16;  
 Matches 34; Conservative 29; Mismatches 51; Indels 25; Gaps 5;  
 QY 51 EDLFLPVKGLKLEGVQPDG-----EISGEGKKDAGVY-----NLKDTFK 95  
 DB 78 EDLFLPVKGLKLEGVQPDG-----EISGEGKKDAGVY-----NLKDTFK 135  
 QY 96 PVFKKIEKKEEENKPT-----FDVSKKQNPQVNHSQLNESHKEDLQREHSGQSDS 149  
 DB 136 PLTELEKLEKEEENKPT-----FDVSKKQNPQVNHSQLNESHKEDLQREHSGQSDS 193  
 QY 150 TKDYATVLDKNNISKST 168  
 DB 194 EGYVATVATTEATTEAT 212  
 RESULT 37  
 ID Q17595 PRELIMINARY; PRT; 535 AA.  
 AC Q17595;  
 DT 01-NOV-1996 (TREMBLrel. 01, Created)  
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)  
 DE Hypothetical 60.3 kDa protein.  
 GN C02H7.1.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
 CC Rhabditidae; Pelodidae; Caenorhabditis.  
 CX NCBI\_TaxID=6239;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RA STRAIN=Bristol N2;  
 RX MEDLINE=99069613; PubMed=9851916;  
 RA None;  
 RT "Genome sequence of the nematode C. elegans: a platform for  
 RT investigating biology. The C. elegans Sequencing Consortium.";  
 RL Science 282:2012-2018 (1998).  
 RN (2)  
 RP SEQUENCE FROM N.A.  
 RA STRAIN=Bristol N2;  
 RX MEDLINE=99069613; PubMed=9851916;  
 RA None;  
 RT "The sequence of C. elegans cosmid C02H7.";  
 RL Submitted (FEB-1996) to the EMBL/Genbank/DBJ databases.  
 RN (3)  
 RP SEQUENCE FROM N.A.

RC	STAIN=Brstcol N2;
RA	"Direct Submission."
RT	Submitted (Aug-2001) to the EMBL/GenBank/DBJ databases.
RL	EMBL; U49945; AAC47924.1; -
DR	Mozmped; C02H7.1; CE06757.
KM	Hypothetical protein.
SO	SEQUENCE 535 AA; 60303 MW; 03B2BBE8B3DFB6 CRC64;

Query Match	11.3%; Score 101.5; DB 5; Length 535;
Best Local Similarity	21.0%; Pred. No. 22;
Matches	39; Conservative 31; Mismatches 71; Indels 45; Gaps 6

Qy	4 VKDFAKNTYVKEFLNK-----DTGEVSLEKPHRYVTWTIONGKMSGVSEDEILPY 58
Dy	55 IKSAGNDKNKTATFADLKLIKILDOSLNANVAAKII-----SKDAKET-----98
Qy	59 KGLLEKGYQDPGWEISPEPKKKDACGYTNLSKDTFKVEPFKIEKKBEENKPTDVASK 118
Dy	99 ----NKKLMQWIGTATNSPNSRNGTG-----EEKKKKKVKEDKCKDEEEKST---TKK 144
Qy	119 KDNPQNVNHSQLESNRKEDLOREHSQKSDSK-----DYTAIVLDKNINSSKST 168
Dy	145 RSSKKEETHEEKESBKXKKAEBEKXKXSSEKRRKKS DRBSSEKSKSEKSKEXKST 204
Qy	169 TNNPNK 174
Dy	205 TDERPK 210

RESULT 38
Q9V719 PRELIMINARY; PRT; 556 AA.
ID Q9V719
AC Q9V719;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-OCT-2002 (TREMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE CG8421 protein.
ASPM OR CG8421 OR CG18658.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
NC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkely;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celitker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoeking R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sulston G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champagne M., Pfeiffer B.D.,
RA Man K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.P., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Bakendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Bernan B.P., Bhattacharya P., Brothier P.,
RA Bookova D., Botchan M.R., Bouck J., Brockstein P., Brotlier P.,
RA Burlis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cuiway S., Dahlke C., Davoport I.B., Davies P.,
de Paaloe B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferrara C., Ferreira S., Fleischmann W.,
RA Feiler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glaser K.,
RA Giodel C., Gong F., Gorrell J.H., Gu Z., Guan P., Harris W.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.Z., Wei M.-H., Idegawa C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Keechum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasok P., Lei Y., Levitsky A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Maceti B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Nobayri C., Morris J., Mosheiff A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Neilson D.L.,

```

RA Nelson D.R., Nelson K.A., Nixon K., Nusken D.R., Pacib J.M.,
RA Pallazzo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Klamos I., Simpson M., Skripski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yen R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zhang Y., Xie H., Zhong P.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
[2]
RN
RP SEQUENCE FROM N.A.
RA Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,
RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
RA Barzon J., An H., Baldwin D., Barzon J., Beeson K.Y., Busam D.A.,
RA Carlson J.W., Center A., Chapple M., Davenport L.B., Dietz S.M.,
RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Fartan D.,
RA Ferreira S., Frishe E., Galle R.F., Garg N.S., George R.A.,
RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
RA Ibegwam C., Jallil M., Kruse D., Li P., Mattei B., Moshrefi A.,
RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
RA Paclib J., Paragas V., Park S., Patel S., Pfeiffer B.,
RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
RA Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,
RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
RT "Sequencing of Drosophila melanogaster genome.";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
[3]
RN
RP SEQUENCE FROM N.A.
RA Miara S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
RA Hirdesky P., Huang Y., Kaminler J.S., Prochuk S.E., Smith C.D.,
RA Turry J.L., Bergman C., Berman B., Carlson J.W., Celniker S.B.,
RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
RA Kronmiller B., Marshall B., Millburn G., Richer J., Russo S.,
RA Searle S.W.J., Smith B., Shu S., Smutnick F., Whitfield E.,
RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
RT "Annotation of Drosophila melanogaster genome.";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
[4]
RN
RP SEQUENCE FROM N.A.
RA Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
[5]
RN
RP SEQUENCE FROM N.A.
RA FlyBase;
RA Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RL EMBL: AB003808; AAF58064.2; -.
DR FlyBase; FBgn0034075; Asph.
SQ SEQUENCE 556 AA; 63144 MW; B420980CDB6C357A CRC64;
Query Match 11.3%; Score 101.5; DB 5; Length 556;
Best Local Similarity 24.5%; Pred. No. 23;
Matches 34; Conservative 29; Mismatches 51; Indels 25; Gaps 5;
Qy 51 EDLPLPYKGELEKGVQFDGW-----EISGFBGKQDAGYIV-----NUSKDTFFIK 95
Db 78 EDDLTPLSESRSFK-VFDGWNDEHHRDHGHVQSPSGALDDHDEHDDHDEDEDBR 135
Qy 96 PVFKLIEKKKEENKPF-----FDVSKKDQNPQVNSQLNESHKRDLDREHSQSDS 149
Db 136 PLTEELERLEERLEERPEFDEBPAADBEYEDBEDENNA--GENVTADEAEERERDND 193
Qy 150 TKQVTAATVLDKNNISSKST 168
Db 194 EGTVEATVETATTAATAT 212
RESULT 39
09G082 PRELIMINARY; PRT; 785 AA.
09G082

```

AC O96082;  
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE Aspartyl beta-hydroxylase variant 1 (CG8421-PA).  
 GN ASH OR CG8421 OR CG18658.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephyridiidae; Drosophilidae; Drosophila.  
 NCBI\_TaxID=7227;  
 RX MEDLINE=20564326; PubMed=10956665;  
 RA Dinchuk J.E., Henderson N.L., Burn T.C., Huber R., Ho S.P., Link J.,  
 RA O'Neill K.T., Pochter R.J., Scully M.S., Hollis J.M., Hollis G.F.,  
 RA Friedman P.A.;  
 RT "Aspartyl beta-Hydroxylase (Asph) and an Evolutionarily Conserved  
 RT Isoform of Asph Missing the Catalytic Domain Share Exons with  
 RT Junction";  
 RL J. Biol. Chem. 275:39543-39554(2000).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Berkley;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blair R.G., Champs M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baker E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abell J.F., Adayant A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Botkova D., Borchan M.R., Bouck J., Brokstein P., Brothier P.,  
 RA Butris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K.J., Doup L.B., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Foster C.J., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kesterson J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Laoko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,  
 RA Palazzolo M., Peltman G.S., Pan S., Pollard J., Puti V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shie B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wasserman D.A., Weissbrock G.M., Weissbrock J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao X., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of Drosophila melanogaster";  
 RL Science 287:2185-2195(2000).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC Celinker S.E., Adams M.D., Krommiller B., Wan K.H., Holt R.A.,  
 RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,  
 RA Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,  
 RA Carlson J.W., Center A., Champs M., Davenport L.B., Dietz S.M.,  
 RA Dodson K., Doseett V., Doup L.B., Doyle C., Dreanek D., Farfan D.,  
 RA Ferreira S., Frise E., Galie R.F., Garg N.S., George R.A.,  
 RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,  
 RA Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,

RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,  
 RA Pacle J., Paragas V., Park S., Patel S., Pfeiffer B.,  
 RA Phouanavong S., Peltman G.S., Puti V., Richards S., Scheeler F.,  
 RA Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,  
 RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;  
 RT "Sequencing of Drosophila melanogaster genome";  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RA Miera S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,  
 RA Hradecky P., Huang Y., Kaminker J.S., Prochnik S.E., Smith C.D.,  
 RA Tupy J.L., Bergman C., Berman D., Frise B., de Grey A., Harris N.,  
 RA Krommiller B., Marshall B., Millburn G., Richter J., Russo S.,  
 RA Searle S.M.J., Smith E., Shu S., Smutnick F., Whitfield B.,  
 RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;  
 RT "Annotation of Drosophila melanogaster genome";  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RA Adams M.D., Celinker S.E., Gibbs R.A., Rubin G.M., Venter J.C.;  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RN [6]  
 RP SEQUENCE FROM N.A.  
 RA FlyBase;  
 RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF289493; AA040806.1; -;  
 DR EMBL; AE003808; AA070947.1; -;  
 DR FlyBase; FBgn0034075; Asph.  
 DR InterPro; IPR001440; TPR.  
 DR InterPro; IPR006025; Zn.MTpeptidse.  
 DR PROSITE; PS00142; ZINC\_PROTEASE; 1.  
 SQ SEQUENCE 785 AA; 89843 MW; 30A8DFCD6836F71 CRC64;  
 Query Match 11.3%; Score 101.5; DB 5; Length 785;  
 Best Local Similarity 24.5%; Pred. No. 34;  
 Matches 34; Conservative 29; Mismatches 51; Indels 25; Gaps 5;  
 QY 51 EDFILPVYKGELEKGYQPDGM-----SIGFGKKGAGYV-----NLKDTFIK 95  
 DB 78 EDLDTPLSESRFSK--VFGWVDEHDEHDDHGDVGPSEGLMDHDDHDDHDEDEDE 135  
 QY 96 PVFKRIEKEKEENKPT-----FDVSKKQNPQVNSQLNSHKKEDLQREHSHQKSS 149  
 DB 136 PLTELEBELEBEKEPTDEBPADBEYEDDEENNA--GNIYAEADBESEEDND 193  
 QY 150 TKDYATATLVLDKNNISKST 168  
 DB 194 EGTVAATVATTEATTEAT 212  
 RESULT 40  
 ID 081556 PRELIMINARY; PRT; 1377 AA.  
 AC 081556;  
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE Eukaryotic translation initiation factor 3 subunit 10, putative.  
 GN Pfl0625C.  
 OS Plasmodium falciparum (isolate 3D7).  
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
 NCBI\_TaxID=36329;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=3D7;  
 RX MEDLINE=22255705; PubMed=12368864;  
 RA Gardiner M.J., Hall N., Fung E., White O., Berriaman M., Hyman R.W.,  
 RA Carlson J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,  
 RA Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,  
 RA Chan M.-S., Nene V., Shalimov S.J., Sub B., Peterson J., Angiuoli S.,  
 RA Perera M., Allen J., Selengut J., Haft D., Mather M.W., Valiya A.B.,  
 RA Martin D.M.A., Fairhead A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Comugen Ltd.

OM protein - protein search, using sw model

Run on: February 10, 2004, 10:48:45 ; Search time 12.5638 Seconds  
(without alignments)  
1331.870 Million cell updates/sec

Title: US-10-067-385-8\_COPY\_600\_773

Perfect score: 897  
Sequence: 1 KIVVDFARNTTKEFIILNK.....ATVLDKNNISSKSTNNPNK 174

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR 76:\*  
1: PIR1:\*  
2: PIR2:\*  
3: PIR3:\*  
4: PIR4:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	897	100.0	2140	2	P95074	serine proteinase,
2	894	99.7	2144	2	A97942	metalloproteinase
3	114.5	12.8	1038	2	UC5497	clausurin - chick
4	110	12.3	558	2	T18467	hypothetical prote
5	109	12.2	665	2	B71609	hypothetical prote
6	104	11.6	540	2	D86432	hypothetical prote
7	103.5	11.5	335	2	T18283	hypothetical prote
8	102.5	11.4	385	2	T20410	hypothetical prote
9	102	11.4	622	2	A90570	lipoprotein [impor
10	101.5	11.3	312	2	G81339	probable membrane
11	101.5	11.3	535	2	T37189	hypothetical prote
12	101	11.3	700	2	S67610	probable membrane
13	100	11.1	219	2	B72291	hypothetical prote
14	100	11.1	1875	2	S38173	myosin-like protei
15	99	11.0	211	2	T25911	hypothetical prote
16	99	11.0	1397	2	T10466	DNA topoisomerase
17	97.5	10.9	614	2	A84152	hypothetical prote
18	97.5	10.9	1345	2	S46817	hypothetical prote
19	97	10.8	644	2	T47835	hypothetical prote
20	97	10.8	988	2	T14188	hypothetical prote
21	96.5	10.8	456	2	T05612	hypothetical prote
22	96.5	10.8	867	2	T27136	hypothetical prote
23	96.5	10.8	871	2	T27135	hypothetical prote
24	96	10.7	491	2	C97267	repeat organellar
25	96	10.7	1939	2	T18372	hypothetical prote
26	95.5	10.6	3724	2	G71609	hypothetical prote
27	95.5	10.6	3724	2	T18427	hypothetical prote
28	95	10.6	348	2	I37271	cylicin II - human
29	95	10.6	1202	1	S05362	probable DNA-direc

30	94.5	10.5	277	2	D70214	surface lipoprotei
31	94.5	10.5	670	2	T28391	ORF MSV230 hypoche
32	94	10.5	210	2	T28771	hypothetical prote
33	94	10.5	553	2	T15094	hypothetical prote
34	94	10.5	947	2	T03795	ornithine decarbox
35	93.5	10.4	456	2	T03045	hypothetical prote
36	93.5	10.4	645	2	B89883	conserved hypothec
37	93.5	10.4	649	2	S42488	dnak-type molecule
38	93	10.4	629	2	G96542	hypothetical prote
39	93	10.4	2464	1	ORMSPI	microtubule-associ
40	92.5	10.3	443	2	S66040	serine-type D-Ala-
41	92.5	10.3	571	2	B64469	hypothetical prote
42	92.5	10.3	1702	2	A41859	Iga-specific metal
43	92.5	10.3	1888	2	T39009	hypothetical prote
44	92.5	10.3	1997	2	F71607	DNA helicase II BR
45	92	10.3	688	2	A47705	triacylglycerol li

## ALIGNMENTS

## RESULT 1

F95074 serine proteinase, subtilase family [imported] - Streptococcus pneumoniae (strain TIGR

C/Species: Streptococcus pneumoniae

C/Date: 03-Aug-2001 #sequence\_revision 03-Aug-2001 #text\_change 03-Aug-2001

C/Accession: F95074

R/Retelin, H.; Nelson, K.B.; Paulsen, I.T.; Bisen, J.A.; Read, T.D.; Peterson, S.; He  
son, J.D.; Unayam, L.A.; White, O.; Salberg, S.L.; Lewis, M.R.; Radune, D.; Holtzaple  
on, T.; Hickey, E.K.; Holt, I.E.  
Science 293, 498-506, 2001

A/Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morriso

A/Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.

A/Reference number: A95000; MUID:21357209; PMID:11463916

A/Accession: F95074

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-2140 <KIP>

A/Cross-References: GB:A8005672; PIRN:NAK74791.1; PID:g14972117; GSPDB:GN00164; TIGR:S

A/Experimental source: strain TIGR4

C/Genetics:

A/Gene: SP0641

Query Match 100.0%; Score 897; DB 2; Length 2140;  
Best Local Similarity 100.0%; Pred. No. 6.9e-57;  
Matches 174; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KIVVDFARNTTKEFIILNKDQGEVSEIKPHRVTTTQNGKMSSTIVSEBDFILPVYK 60

DB 1933 KIVVDFARNTTKEFIILNKDQGEVSEIKPHRVTTTQNGKMSSTIVSEBDFILPVYK 1992

QY 61 ELEKGYQPDGWEISGPEKGDAGYVNIISKOTFIKPVFKIEEKEENKPTFDVSKKD 120

DB 1993 ELEKGYQPDGWEISGPEKGDAGYVNIISKOTFIKPVFKIEEKEENKPTFDVSKKD 2052

QY 121 NPQVNSQLNESHKREDLQREHSHQKSDSTKQVTATVLDKNNISSKSTNNPNK 174

DB 2053 NPQVNSQLNESHKREDLQREHSHQKSDSTKQVTATVLDKNNISSKSTNNPNK 2106

## RESULT 2

A97942 metalloproteinase (EC 3.4.21.-) A [imported] - Streptococcus pneumoniae (strain R6)

C/Species: Streptococcus pneumoniae

C/Date: 22-Oct-2001 #sequence\_revision 22-Oct-2001 #text\_change 02-Nov-2001

C/Accession: A97942

R/Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; Dehoff, B.S.;  
e, R.; Leblanc, D.J.; Lee, L.N.; Letkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.;  
Y, P.; Sun, P.M.; Winkler, M.B.  
J. Bacteriol. 183, 5709-5717, 2001

A/Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaekunas, S.I

A/Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.

A/Reference number: A97872; MUID:21429245; PMID:11544234

A:Accession: A97942  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-2144 <KTR>  
A:Cross-references: GB:AE007317, PIDN:AAK9365.1, PID:G1548138, GSPDB:GN00174  
C:Genetics:  
A:Gene: pTCA  
C:Keywords: hydrolase; serine proteinase

Query Match 99.7%; Score 894; DB 2; Length 2144;  
Best Local Similarity 99.4%; Pred. No. 1.1e-56;  
Matches 173; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 KIVVQDPARNTVKEFIILNKDTGSEVSELPKPHRYVTITIQNGKMSSTIVSEEDFILPVYKG 60  
DB 1937 KIVVQDPARNTVKEFIILNKDTGSEVSELPKPHRYVTITIQNGKMSSTIVSEEDFILPVYKG 1996  
QY 61 ELEKGYQPDGWEISGFEKGKQAGYVNLTKDTPFKVPEKKIEKKKEENKPFEDYSKKD 120  
DB 1997 ELEKGYQPDGWEISGFEKGKQAGYVNLTKDTPFKVPEKKIEKKKEENKPFEDYSKKD 2056  
QY 121 NPQVNSHQLNSHREKDLQREDSQKSDSTKVATVTLKNNISKSTNNPK 174  
DB 2057 NPQVNSHQLNSHREKDLQREDSQKSDSTKVATVTLKNNISKSTNNPK 2110

## RESULT 3

JCS497  
N:Alternate names: keratan sulfate proteoglycan  
C:Species: Gallus gallus (chicken)  
C:Date: 07-Jul-1997 #sequence revision 12-Sep-1997 #text\_change 21-Jul-2000  
C:Accession: JCS497; PC4334; S37561  
R:Burg, M.A.; Cole, G.J.  
J. Neurobiol. 25, 1-22, 1994  
A:Title: Claustrin, an antiadhesive neural keratan sulfate proteoglycan, is structurally  
A:Reference number: JCS497; MUID:94157526; PMID:7906711  
A:Accession: JCS497  
A:Molecule type: mRNA  
A:Residues: 1-1038 <BUR1>  
A:Cross-references: EMBL:X67778; NID:G406318; PIDN:CAA47988.1; PID:G406319  
A:Accession: PC4334  
A:Molecule type: Protein  
A:Residues: 79-83;299-412;485-502 <BUR2>  
A:Experimental source: brain  
C:Comment: This protein inhibits neural cell adhesion and neurite outgrowth in the nervous system.  
F:267-270/Region: cell attachment (R-G-D) motif  
F:112,213,450/Binding site: carbohydrate (Asn) (covalent) #status predicted  
F:152,249,440,793,820/Binding site: chondroitin sulfate (Ser) (covalent) #status predicted

Query Match 12.8%; Score 114.5; DB 2; Length 1038;  
Best Local Similarity 23.5%; Pred. No. 1.3;  
Matches 47; Conservative 29; Mismatches 65; Indels 59; Gaps 6;

QY 2 IYVQDPARNTVKEFIILNKDTGSEVSELPKPHRYVTITIQNGKMSSTIVSEEDFILPVYKG 39  
DB 514 IYVQDPARNTVKEFIILNKDTGSEVSELPKPHRYVTITIQNGKMSSTIVSEEDFILPVYKG 573  
QY 40 -----GKMSSTIVSEEDFILPVYKGELEKGYQPDGWEISGFEKG 79  
DB 574 LEKQGLSEKERTPVKKEKAVKPEYTYIAEKV-----TTKEBQGLSKSTSKQAS 625  
QY 80 KQAGYVNLTKDTPFKVPEKKIEKKKEENKPFEDYSKKDNPQVNSHQLNSHREKDL 138  
DB 626 EKQDQVPEKTKKSVKAKPEKKDKKPKKSVKKEKPEL-----KKEKPKKEDI 662  
QY 139 QRE-----SHQKSDSTKV 153  
DB 683 KKEVKEVKEVKEVKEVKEV 702

## RESULT 4

T18467  
hypothetical protein C0465C - malaria parasite (Plasmodium falciparum)  
C:Species: Plasmodium falciparum  
C:Date: 15-Oct-1999 #sequence revision 15-Oct-1999 #text\_change 11-Jan-2002  
C:Accession: T18467  
R:Lawson, D.; Bowman, S.; Barrell, B.  
submitted to the EMBL Data Library, November 1998  
A:Reference number: Z18937  
A:Accession: T18467  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-558 <LAW>  
A:Cross-references: EMBL:AL008970; NID:el407852; PIDN:CAA15610.1

Query Match 12.3%; Score 110; DB 2; Length 558;  
Best Local Similarity 29.3%; Pred. No. 1.3;  
Matches 54; Conservative 23; Mismatches 47; Indels 60; Gaps 12;

QY 47 IYSEEDFILPVY-----KGELEKGYQPDGWEISGFEKGK-----DAGYVNLTKDTPFKV 97  
DB 60 ILGFEDDILVEYICISQLKQSKK--KADSEEDKYLNAKTLKNTLGFIGNKKSDFIEEL 117  
QY 98 FKXI--EEKKEE-----ENKPFEDVSK-KQDNPOVNSHQLNSHREKDLQREDSQKSDSTKVATVTLKNNISKSTNNPK 135  
DB 118 LELLNEEKKEEHIADTLNENK--TNDIKKVKNNENINENYNNENKDNISNKEKSHQON 176  
QY 136 E-----DLQREH-----SOKSDSTK---DVTATVTLKNNISKSTNNPK 170  
DB 177 EHNINNVNLKKEKEYTDIQRDKRKRRSLQKSDSTKVATVTLKNNISKSTNNPK 235  
QY 171 NPNK 174  
DB 236 KTNK 239

## RESULT 5

B71609  
hypothetical protein PFB0680W - malaria parasite (Plasmodium falciparum)  
C:Species: Plasmodium falciparum  
C:Date: 13-Nov-1998 #sequence revision 13-Nov-1998 #text\_change 21-Jul-2000  
C:Accession: B71609  
R:Gardner, M.J.; Tetteh, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.; Pettes, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.C.  
Science 282, 1126-1132, 1998  
A:Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.  
A:Reference number: A71600; MUID:99021743; PMID:9804551  
A:Accession: B71609  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-665 <GAR>  
A:Cross-references: GB:AE001410; GB:AE001362; NID:G3845245; PIDN:AACT1925.1; PID:G3845;  
A:Experimental source: clone 3D7  
C:Genetics:  
A:Gene: PFB0680W

Query Match 12.2%; Score 109; DB 2; Length 665;  
Best Local Similarity 26.1%; Pred. No. 1.9;  
Matches 49; Conservative 32; Mismatches 49; Indels 58; Gaps 11;

QY 5 KQFARNTVKEFIILNKDTGSEVSELPKPHRYVTITIQNGKMSSTIVSEEDFILPVYGELE 63  
DB 114 KQNNNNNGTKQIBKKNKINKSDL--HQNBLNLQSGK-----NKNE 158  
QY 64 KGYQPDGWEI--SGFEKGKQAGYVNLTKDTPFKVPEKKIEKKKE-----ENKPK 111  
DB 159 KGMQ-----DISNSNENKQD-----VKEGVKELEKKEKKEKISDDHKVBNKK 202  
QY 112 TFD-----VSKQDNPOVNSHQLNSHREKDLQREDSQKSDSTKVATVTLKNNISK 166

Db 203 SDDHKVEENKSDHAKVEENKSDHAKIEVKVEEHEDDEE-----DKKKKSE 253  
QY 167 STTNPNK 174  
Db 254 MNKMDENK 261

## RESULT 6

D66432

hypothetical protein T518.14 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Dec-2001

C:Accession: D66432

R:Theologos, A.; Ecker, J.R.; Palm, C.J.; Pederspiel, N.A.; Kaul, S.; White, O.; Alonso,

Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;

ansen, N.F.; Hughes, B.; Huizar, L.

Nature 408, 816-820, 2000

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.

C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maltl, R.; Marshall,

Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A:Authors: Salazar, S.L.; Schwartz, J.R.; Shim, P.; Southwick, A.M.; Sun, H.; Tallon,

ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A:Reference number: A66141; MUID:21016719; PMID:11130712

A:Accession: D66432

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-540 &lt;STO&gt;

A:Cross-references: GB:AB005172; NID:g4587525; PIDN:AAD25756.1; GSPDB:GN00141

C:Genetics:

A:Map position: 1

Query Match 11.6%; Score 104; DB 2; Length 540;

Best Local Similarity 25.3%; Pred. No. 3.5; Mismatches 64; Indels 26; Gaps 6;

Matches 40; Conservative 28; Mismatches 64; Indels 26; Gaps 6;

QY 11 TVKEPILNKDNGEVELKPHRYVTITONGKEMSTIVSEDFILPYKGELEKGYQFDG 70

Db 27 TVKKA-VSEETKKEBESRP-----EGVEKSAKPKESDFFADLKESKK----- 70

QY 71 WEISGFEKKDAGYVILNSKDTPIKPVFKK--IEKKKEENKPTDVSCKKDPQVNH 127

Db 71 -ALSDKSLBEATVON---TLKTKKSSPMKEKKEVAVKPAVEKKK--BAAE 123

QY 128 QLNESHKEDLOREHSQKSDSTQDVTATVLDNNISS 165

Db 124 KYBEKKSAVVTBAPKATVAVVTEIIPKEBVT 161

RESULT 7  
T18283  
hypothetical protein GS - alime mold (Dictyostelium discoideum)  
C:Species: Dictyostelium discoideum  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
C:Accession: T18283  
R:Reben Jr., W.K.; Gonzales, C.M.; Gonzales, S.T.; Pilkington, K.J.; Kiyosawa, H.; Hugi-

Genetics 148, 1117-1125, 1998

A:Title: Dictyostelium discoideum nuclear plasmid Ddps is a chimera related to the Ddpl

A:Reference number: Z14684; MUID:9819836; PMID:9539429

A:Accession: T18283

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-325 &lt;RIB&gt;

A:Cross-references: EMBL:U00796; NID:g2702254; PID:g2702258; PIDN:AA018634.1

C:Genetics:

A:introns: 85/1

Query Match 11.5%; Score 103.5; DB 2; Length 325;

Best Local Similarity 23.8%; Pred. No. 2.1;

Matches 43; Conservative 25; Mismatches 50; Indels 63; Gaps 9;

QY 31 HRYVTITONGKEMSTIVSEDFILPYK-GELEKGYQFDGMEISGFEKK----- 79

Db 31 HRYVTITONGKEMSTIVSEDFILPYK-GELEKGYQFDGMEISGFEKK----- 79

QY 31 HRYVTITONGKEMSTIVSEDFILPYK-GELEKGYQFDGMEISGFEKK----- 79

Db 31 HRYVTITONGKEMSTIVSEDFILPYK-GELEKGYQFDGMEISGFEKK----- 79

QY 31 HRYVTITONGKEMSTIVSEDFILPYK-GELEKGYQFDGMEISGFEKK----- 79

Db 31 HRYVTITONGKEMSTIVSEDFILPYK-GELEKGYQFDGMEISGFEKK----- 79

QY 31 HRYVTITONGKEMSTIVSEDFILPYK-GELEKGYQFDGMEISGFEKK----- 79

Db 31 HRYVTITONGKEMSTIVSEDFILPYK-GELEKGYQFDGMEISGFEKK----- 79

QY 31 HRYVTITONGKEMSTIVSEDFILPYK-GELEKGYQFDGMEISGFEKK----- 79

Db 31 HRYVTITONGKEMSTIVSEDFILPYK-GELEKGYQFDGMEISGFEKK----- 79

QY 31 HRYVTITONGKEMSTIVSEDFILPYK-GELEKGYQFDGMEISGFEKK----- 79

Db 31 HRYVTITONGKEMSTIVSEDFILPYK-GELEKGYQFDGMEISGFEKK----- 79

QY 31 HRYVTITONGKEMSTIVSEDFILPYK-GELEKGYQFDGMEISGFEKK----- 79

Db 31 HRYVTITONGKEMSTIVSEDFILPYK-GELEKGYQFDGMEISGFEKK----- 79

QY 31 HRYVTITONGKEMSTIVSEDFILPYK-GELEKGYQFDGMEISGFEKK----- 79

Db 31 HRYVTITONGKEMSTIVSEDFILPYK-GELEKGYQFDGMEISGFEKK----- 79



A;Gene: MYPU\_4650  
A;Genetic code: SGC3

Query Match 11.4%; Score 102; DB 2; Length 622;  
 Best Local Similarity 24.9%; Pred. No. 5.7;  
 Matches 48; Conservative 35; Mismatches 66; Indels 44; Gaps 11;

```

Qy      9 RNTTYKEFIL-NKDTGEVSEL-----KPRHV--TYVTIONG--KEM--STI--V 48
      10 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      26 KNDASAKNFVLSKSSDLISISEKQFKPKRNLDKQKQKIEGVITFINSETYKRIVKKETTILNL 85

```

Qy	49	SEEEPILPVKKGELEKGYOFPGMEISGFEKKDAGYINLSKDTFIPVFKLIEEKKEE	108
Db	86	SEDNIIFSILNTEIENNAKFOLDEF-----VSADKEKFIKFGSINFSGTQQ	129

QY 109 NKPTEDVSKKKD---NPQVHNSQLNHS--HKREDIQRERHSOKSDSDTQVATATVLKGN- 161  
 130 -KIDNITSSKEDEKNGKPKDENNSNNSSDQKDELQKNSGSKDNDVQDEKANKENSNS 188

QY	162	NISSKSTNNPNK	174
		:	
Db	189	NDSKEKDENTNK	201

## RESULT 10

```

probable_mammaprotein Cj0692c [imported] - Campylobacter jejuni (strain NCTC 1168)
C.Species: Campylobacter jejuni
C.Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 03-Jun-2002

```

R. Parhill, M. W. Munneil, K. J. Kelley, C. J. Basham, D. J. Chaffler, C. M. Rajandream, M. A. Rutherford, K. M. VanVleet, A. Whitehead, S. Barrett, 665-668/2000

A:Reference number: AB1250; MUID:20150912; PMID:10688204  
A:Accession: G91339  
A:Status: preliminary

A:Residues: 1-312 <PAR>  
A:Cross-references: GB:AL131168, NID:g59668128; PIDN:CAB72366.1; PID:g5968128  
A:Experimental source: serotype O2, strain NCTC 11168

Query Match	11.3%	Score 101.5	DB 2	Length 312
A:Gene: Cj0692c				

15	EF1LNDTGEVSELKPRHTVTIIONGKMSSTI	VS	BBDFILEPVK	-----	GHLE	63
	Matches	46;	Conservative	34;	Mismatches	73;
					Indels	31;
					Gaps	9;

Db 5 DFIKERONIKOMLKEFSR---AINOSKPELDDIRDEIRSSDILIRKRFKCKTPNKLEELD 61  
Oy 64 KGYOFDGMELISGFESKKDAGYVNL---SKOTFIKVFKKLEBKKEBENKPT---PDVSK 117

Db 62 EYESHAKTKSNITLYKED--LINVLEBKOSLAKI FSKOKERKEKENTKTAKKFLPSR 118

Qy 118 KKNP---QVNSQLNSHRKEDLOREHSOKSDTKV--TATVLDKNNISK--STT 169

Db	119	KKANEIKNIOTKTQIQTKSNQAITQTKQEKELTNSIEKIQTEIKIQKPLIEKKLDVK	178
Qy	170	NNPN	173

Db 179 NOFN 182

hypothetici

C:\Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 18-Feb-2000  
C:\Accession: T37189  
R.\leimbac, D.; Minx, M.  
Submitted to the EMBL Data Library, February 1996

A;Description: The sequence of *C. elegans* cosmid C02H17.  
A;Reference number: Z20523

A/Accession: J27105  
A/Status: preliminary; translated from GB/EMBL/DBJ  
A/Molecule type: DNA  
A/Residues: 1-535 <LEI>

A:Accession: AF009333, GenBank  
A:Experimental source: strain Bristol N2; clone C02H7  
C:Genetics:  
A:Gene: CESP:C02H7.1

Query Match	11.3%	Score 101.5	DB 2	Length 535
Align Position: A				
A: Introns: 47/3;	100/3;	149/3;	304/2;	347/3; 458/3

QY 4 VKDPARTTVEFLINK-----DTGEVSELKPHRYVTIQLNGKMSSTIVSEEDFILPVY 56

```

Db      55  IKSAGNDKNTTAFLDRLIKILDDGLNKNVKAALI---SGDAEET----- 98
QY      59  KGELEKGYQFDGWEISGPEGKDXAGYVINTLSKDTPIKPYFKIIEKKEBENKPTPDVSKK 11

```

Db	99	---NKLQMLGTNATSFNSRNGTG-----	EEKCKKKVKKEDPKGDEEEKST---	TKK	14
Qy	119	KDNPQVNHSQLNESHKEDLDREHSHQKSDTK-----	DYATATVLDKNNISKST		16

Db 145 RSSKGETHEEKSEKSSAEEKKKKSSSSKRRHKSDDRSEKSSBKSSKKKEKST 20

RESULT 12  
S67610  
probable membrane protein YDL074c - yeast (*Saccharomyces cerevisiae*)

C:Accession: S67610  
 C:Date: 12-Jul-1996 #sequence\_revision 12-Jul-1996 #text\_change 19-Apr-2002  
 C:Species: *Saccharomyces cerevisiae*  
 N:AlternativeNames: hypochelate protein 220S

A;Accession: S67610

A: molecule type: DNA  
A: Residues: 1-700 <NAM>  
A: Cross-references: EMBL: Z74122, NID: G1431087, PID: e253213, GSPDB: GN0000  
A: Experimental source: strain S288C

C;Genetics:  
A;Gene: SGD:BRE1, MIPS:YDL074C  
A;Cross-references: SGD:S0002232  
A;Map position: 4L

Query Match 11.3%; Score 101; DB 2;  
C/keywords: transmembrane protein  
F;69-85/Domain: transmembrane #status predicted <TMM>

Best Local Similarity	22.4%;	Freq. no. / . / ?
Matches	46;	Conservative 39; Mismatches 60; Indels 60; Gaps 1
5 KDFARNTTVKSEILNK-----DNGEVSSELKPHRYVTITQ-----NGKEMSSST 46		

```

Db      379  KDVARIRTRADLLSKIAILEAEKSTEVLSDLQ-HAIDLKEQWTKIDKRSNDTKSSST 433
QY      47   IVSEEDFILPVYKGELEKGYQPDGWEISGFEGK-----DAGYV 85

```

```

Db      438  ---QDALKI:Q-DLEKGR---ELSDLTHKYSEIINHESVLSKLTVEKTKADQKYF 48
QY      86  INL-SKDTF---IKPVFKKILEEKKEEKKPTFDVSKKKKNPQVHNSQLNESHKREDIQRE 14

```

```

Db      489 AAMRSKDSILTEIKTSLSKSLSKNEL-----ILQKSDRLLOQKIGNLHKQLDLSQN 544
QY      142 EHSQKSDSTKQVATATVLDKNNISSK 166

```

Db 542 NERRLIDSSKTEFLTKIIDLNNTSTK 566

RESULT 13  
B72291  
hypothetical protein - Thermotoga maritima (strain MS8)  
C:Species: Thermotoga maritima  
C:Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 21-Jul-2000  
C:Accession: B72291  
R:Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gilm, M.L.; Dodson, R.J.; Haft, D.H.; Hickey,  
Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.  
C.M.

Nature 399, 323-329, 1999  
A>Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome seq  
A:Reference number: A72200; PMID:99287316; PMID:10360571  
A:Accession: B72291  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-219 <ARN>  
A:Cross-references: GB:AE001771; GB:AE000512; MID:g4981678; PIDN:AAD36218.1; PID:g498168  
A:Experimental source: strain MS8  
C:Genetics:  
A:Gene: TM1142

Query Match 11.1%; Score 100; DB 2; Length 219;  
Best Local Similarity 26.4%; Pred. No. 2.4;  
Matches 47; Conservative 30; Mismatches 59; Indels 42; Gaps 10;

Oy | NTTVKEPLANKDGEVSELEKPHRVTVTQNGKMSSTIVSEDFILPVYKGELKGYO-- 67  
| :  
Db | LNFKKKEFPFKEDFOEI--LIPNFVVELYATDEKTTLPKKE-----VLGSESYSVRDL 98  
| :  
Oy | 68 PDGMEISG-----FEGRKDAGYVIN-LSKDTFIRVPFKIEEKENKPFDVSKKD 120  
| :  
Db | 99 FAFGVARGVTTPFFPKSGEGLYLGVDKNFIK-LKYVAQLKED----FGTLAKD 153  
| :  
Oy | 121 NPQVNSHQLNESHKED---LQREHSQKSQS-----TYDVATATVLDK 160  
| :  
Db | 154 DPVVGPELIIRIF-KEDMDPVLKEDENAVKVDTVNEVARDRITVTDSPDAKTLDK 210  
| :

RESULT 14  
S38173  
myosin-like protein MLPI - Yeast (Saccharomyces cerevisiae)  
N:Alternate names: protein YKR095w; protein YKR415  
C:Species: Saccharomyces cerevisiae  
C:Date: 03-May-1994 #sequence\_revision 03-May-1994 #text\_change 21-Jul-2000  
C:Accession: S38173; S40647; S31207  
R:Baldron, V.; Ballesta, J.P.G.; Bou, G.; del Rey, F.; Esteban, P.F.; Garcia-Cantalejo,  
submitted to the Protein Sequence Database, March 1994  
A:Reference number: S38158  
A:Accession: S38173  
A:Molecule type: DNA  
A:Residues: 1-1875 <BAL>  
A:Cross-references: EMBL:D29320; MID:g486586; PID:g486587; MIPS:YKR095w  
A:Experimental source: strain S288C  
R:Kou, G.; Esteban, P.F.; Baldron, V.; Gonzalez, G.A.; Cantalejo, J.G.; Remacha, M.; J  
Yeast 9, 1349-1354, 1993  
A>Title: The complete sequence of a 15 820 bp segment of Saccharomyces cerevisiae chromo  
A:Reference number: S40644; MID:94205265; PMID:8154186  
A:Accession: S40647  
A:Molecule type: DNA  
A:Residues: 1-1875 <BOU>  
A:Cross-references: EMBL:X75541; MID:g450550; PIDN:CAA51948.1; PID:g450554  
A:Experimental source: Strain S288C  
R:Koelling, R.; Nguyen, T.; Chen, B.Y.; Botstein, D.  
Mol. Gen. Genet. 237, 359-369, 1993  
A>Title: A new yeast gene with a myosin-like heptad repeat structure.  
A:Reference number: S31207; MID:93247549; PMID:8483450  
A:Accession: S31207  
A:Molecule type: DNA  
A:Residues: 1-300; 'A', 302-1875 <KOB>

C:Genetics:  
A:Cross-references: EMBL:L01992; MID:g171958; PIDN:AAA4/BJ.L.; PLU:g1/1959  
A:Gene: SGD:MUP1  
A:Cross-references: SGD:S0001803; MIPS:YKR095W  
A:Map position: 11R

Query Match      11.1%; Score 100; DB 2; Length 1875;  
Best Local Similarity    24.4%; Pred. No. 26;  
Matches         50; Conservative    38; Mismatches    71; Indels    46; Gaps    10;

Oy                  4 VKDFARNTTVKPFILNKDTGVEVSELPHRVTTVTIONGKMSSSTIVSEEDF-----ILPVY 58  
| | : | : | : | : | : | : | : | : | : | : | : | :  
Db                 647 ISQITRESTETKNLSLNK---EIQDLYDKSDISIKLGKEKKSRILLAEERFKLLSNTLDLT 703  
  
Oy                 59 KGE--LEKGQFDMEISGFEGCKDA-----GYINLSKDPTFLPKPFIIEK---- 104  
| | : | : | : | : | : | : | : | : | : | : | : | :  
Db                 704 KAENDQLKKRPDY----LONTILKODSKHETHLAENVSCSKSLIYTETELMLNEEKDLR 759  
  
Oy                 105 -----XBEKKPPFPDVSKKDNPOVNHSOLNESHR-REDL---QREHSOKDSSTDGV 153  
| | : | : | : | : | : | : | : | : | : | : | : | :  
Db                 760 VHLERNLKQDELNRK---LSPBKDSLRIWTQTLOTQRKERDELLLETTRKSCQQKIIDLEDA 815  
  
Oy                 154 TA-----TVDKNNISKSSTNNPN 173  
| : | : | : | : | : | : ||||  
Db                 816 LSELAKETSQKDHIIKOLEDENNNSN 840

RESULT 15

T25911  
hypothetical protein T23B3.5 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
C:Accession: T25911  
R.Maggi, B.J. Le, T.  
submitted to the EMBL Data Library, February 1997  
A:Description: The sequence of C. elegans cosmid T23B3.  
A:Reference number: Z20109  
A:Accession: T25911  
A>Status: preliminary; translated from GB/EMBL/DDbfJ  
A:Molecule type: DNA  
A:Residues: 1-211 <MAG>  
A:Cross-references: EMBO:UB8309; PIDD:AAB4234.1; GSPPDB:GN00019; CBSP:T23B3.5  
A:Experimental source: strain Bristol N2; clone T23B3  
C:Genetics:  
A:Gene: CBSP:T23B3.5  
A:Map position: 1  
A:Introns: 30/2; 200/3

Query Match      11.0%; Score 99; DB 2; Length 211;  
Best Local Similarity    29.4%; Pred. No. 2.7;  
Matches         30; Conservative    17; Mismatches    51; Indels    4; Gaps    1;

Oy                  77 EGKADAGVIYNLSKTFFIPVPFKTIIEKKKEEEKKPFTPDVSXXXXNPQVNH-----SOLNES 132  
| | : | : | : | : | : | : | : | : | : | : | : | :  
Db                 69 EGEEKKDGKKRKDDKKBBEKKDGKGGDKEDDKDEDKDXADEKKDE 128  
  
Oy                 133 HKREDQREHHSGXSDTDVATATVLIDKNINSKSTTNPNK 174  
| | : | : | : | : | : | : | : | : | : | : | : | :  
Db                 129 EKDDKDRKBKDKKDEKDEKKESKSKSKSKSKSKSKSKSKSK 170

RESULT 16

T10466  
DNA topoisomerase (ATP-hydrolyzing) (EC 5.99.1.3) II - malaria parasite (Plasmodium fa  
C:Species: Plasmodium falciparum  
C:date: 16-Jul-1999 #sequence\_revision 16-Jul-1999 #text\_change 20-Jun-2000  
C:Accession: T10466  
R.Cheeseman, S.J.  
submitted to the EMBL Data Library, September 1995  
A:Reference number: Z17031  
A:Accession: T10466  
A>Status: preliminary; translated from GB/EMBL/DDBfJ  
A:Molecule type: DNA





[illegible][illegible]

```

Query Match      10.7%; Score 96; DB 2; Length 2500;
Best Local Similarity 26.1%; Pred. No. 77;
Matches         41; Conservative 30; Mismatches 54; Indels 32; Gaps 7;

QY      29 KPHRVVLTONGKMSSTIVSEDFILPVYKGELGYQDGMETISGPEGKKDAGYVINTL 88
        ||:::||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::
Db       2173 KPDKIT--ENNKK-----NEGNEILLKYSIENEHKNYNDKGENECILDKDQGVNNT 2223

QY      89 -----SKDTFIKPVFKKILEEKKEEENKPTFVVSKKKQNPQVNSQLNE---SHRKED 137
        :||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::
Db       2224 KEKNLIDNKKSFPFSNIKYVLEBEESKDDKRD---DKKNDVTREKNLIDNKKSFPFSNIKVK 2280

QY      138 LQREHSOKSDSTKDVATVFLDKNNISSKSTTNPNK 174
        ::||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::
Db       2281 IREER--RSDPKRD-----DKKNDVTREKNLIDNK 2308

```

[illegible]

```

Qy      146 KSDSTKDVATVLDKNNISSKSTNNPNK 174
          | : : ::||: || |
Db      1137 KGNDM-----NENNMENKSDIENENK 1158

```

**RESULT 28**

Cylyictin II - human  
C:Species: Homo sapiens (man)  
C:Date: 12-Aug-1996 #sequence\_revision 12-Aug-1996 #text\_change 21-Jul-2000  
C:Accession: I37271; S52774  
R:Hees, H.; Held, H.; Zimbelmann, R.; Franke, W.W.  
Exp. Cell Res. 218, 174-182, 1995  
A:Title: The protein complexity of the cytoskeleton of bovine and human sperm heads: the  
A:Reference number: I37271; M0ID:95255491; EPIID:7737358  
A:Accession: I37271  
A:Status: preliminary; translated from GB/EMBL/DDBJ  
A:Staccus: preliminary; mRNA  
A:Molecule type: mRNA  
A:Residues: 1-348 <HRS>  
A:Cross-references: EMBL:246788; NID:g758586; PIDN:CA06752.1; PID:g758587

Query Match	10.6%;	Score 95;	DB 2;	Length 348;
Best Local Similarity	30.9%;	Pred. No. 9.4;		
Matches	38;	Conservative	20;	Mismatches 39;
				Indels 26;
				Gaps 7;

[illegible]

QY	164	SSK	166
Db	314	DSK	316

## RESULT 29

C:Probable DNA-directed DNA polymerase (EC 2.7.7.7) - fungus (Ascoclolus immerus) mitoch  
 C:Species: mitochondrion Ascoclolus immerus  
 C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 16-Jun-2000  
 C:Accession: S05362  
 R:Kempken, F.; Meinhardt, F.; Esser, K.  
 Mol. Gen. Genet. 218, 523-530, 1989  
 A>Title: In oranello replication and viral affinity of linear, extrachromosomal DNA of  
 A:Reference number: S05362; M0ID:90066356; PMID:2573821  
 A:Accession: S05362  
 A:Molecule type: DNA  
 A:Residues: 1-1202 <KEM>  
 A:Cross-references: EMBL:X15982; NID:G2933; PIDN:CNA34106.1; PID:G1370212  
 C:Genetics:  
 A:Genome: mitochondrion  
 A:Genetic code: SGC3  
 C:Superfamily: Ascoclolus probable DNA-directed DNA polymerase  
 ;Keywords: DNA binding; mitochondrion; nucleotidyltransferase

Query Match	10.6%;	Score 95;	DB 1;	Length 1202;
Best Local Similarity	22.8%;	Pred. No. 39;		
Matches 44;	Conservative 37;	Mismatches 80;	Indels 32;	Gaps 9;

```

QY      10 NTYVAFILAK----DTGEVSEIKPHRTVTIONGKEMSTI---VSEELP--ILPYVK 60
      |||:::|||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      305 NPTGPFFVNVAKIKPPFGNVRISIGFGNT--TLTDKETILIKTLAIFLEREDITHVMSYDEG 366

```

Db QY 61 ELER-----GYFGWMEISGFBGKDGAGYINLSKDTFIKEVFKIE-----EKKEE 108KKE

364 DIDBSKFPKKSLSFDPKLNTIBGTCYANYTFPIKDIIVAMDINKKINFNGLDLPKRTMDL 423KSK

Qy	109	NK	-	TFDVS	KKKD	NOV	NSH	QRE	HS	OKS	DS	TK	OVAT	VL	DKNI	----	163								
	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:								
Db	424	SKWNL	KL	KND	KT	SGE	I	RM	TK	KN	OKS	YDI	-	IGH	I	IND	EN	VT	IF	RA	VD	NS	IL	KXE	4811

QY	164	---	SSKSTYNNPN	173
Db	482		TVTDSMGNTNDPN	494

RESULT 30

surface lipopeptide p27 homolog - Lyme disease spirochete plasmid A/lps4  
C.Species: Borrelia burgdorferi (Lyme disease spirochete)  
C.Date: 13-Feb-1998 #sequence\_revision 13-Feb-1998 #text\_change 08-Oct-1999  
C.Accession: D70214  
R.Fisser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; Whitt  
son, D.; Peterson, J.; Krelavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vy  
s, Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.  
Nature 390, 580-586, 1997  
A.Authors: Smith, H.O.; Venter, J.C.  
A.Title: Genomic sequence of a Lyme disease spirochete, Borrelia burgdorferi.  
A.Reference number: A70100; MUID:98065943; PMID:9403685  
A.Accession: D70214  
A.Status: preliminary; nucleic acid sequence not shown; translation not shown  
A.Molecule type: DNA  
A.Residues: 1-277 <KLE>  
A.Cross-references: GB:A000790; NID:g2690224; PIDN:AAC66239.1; PID:g2690238; TIGR:BB7  
A.Experimental source: strain B31  
C.Genetics:  
C.Genome: plasmid

Query Match	10.5%	Score 94.5	DB 2	Length 277
Best Local Similarity	20.9%	Pred. No. 7.8		
Matches 34, Conservative	28	Mismatches 62	Indels 39	Gaps 4

Qy	4	VKDFAKNTTVEFLINKQGVSEKPHVYVYI	QNGKMSSTI	VSABD	ILVYV	QGLE	63
		:: :: :: :: :: :: :: :: :: :: :: :: ::					
Db	142	LKNFEIKNIENIFQNDLFLVLTLDKNNNTIN	IMLPNDI	QKPKYIL	LDL	KDTIK	201
		:: :: :: :: :: :: :: :: :: :: :: :: ::					
Qy	64	KGYQFDGWEISGFSKQDAGVYINISKQFIE	KPVFEKI	EKKKEENK	PF	DSVKDQNDPQ	123
		:: :: :: :: :: :: :: :: :: :: :: :: ::					
Db	202	KG-----TSEK-----YLPYR-----				FOIKQKUDYHS	225
		:: :: :: :: :: :: :: :: :: :: :: :: ::					

```
OY      124 VNH$QLNESHKREDLQREHSHQSXSDSTKVTA---TVLDKNKI   163
          ::::: | : | | | : | : | : ||| : 
Db     226 IDYNKVITISEKTIELDLLPHEQVFQMNNKNFTKILDTITDLNL    268
```

**RESULT 31**

ORF MSV230 hypothetical protein - Melanoplus sanguinipes entomopoxvirus  
C:Species: Melanoplus sanguinipes entomopoxvirus  
C:Date: 21-Jan-2000 #sequence\_revision 21-Jan-2000 #text\_change 21-Jul-2000  
C:Accession: J78391  
J:Alonso, C.L., Taiman, E.R., Lu, Z., Oma, E., Kutish, G.F.: Rock, D.L.  
J. Virol. 73, 533-552, 1999  
A:Title: The genome of Melanoplus sanguinipes entomopoxvirus.  
A:Reference number: 220484; MUID:99102612; PMID:5847359  
A:Accession: J78391  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-670 <AFO>  
A:Cross-references: EMBL:AF063866; NID:g4049647; PIND:AA097713.1; PID:g4049753  
C:Genetics:  
A:Note: MSV230

Query Match	10.54	Score	94.5	DB	2	Length	670
Best Local Similarity	21.94	Pred. No.	22				
Matches	42	Conservative	39	Mismatches	82	Indels	29
						Gaps	6

```

QY      3 VVKDEARNTYKEFLINKDTGEVSELEPHRYTYIUNGKEMSSFI-----VS 49
      :: | ::|| | | : ||: ||
Db      9 ILDEKVINMKIEKFI-----KLAKYKYDYNVFEKKQIISILLKFNPFDKTEMCGVS 60

```

```

04      50 EEDFIFPVYKGELEKGY-----QFDGWEISGFEGKGDAGYVINSKOTFIPKVFKE-E 103
      | | : : : | : | : | : | : | : | : | : | : | : | : | : | : | : |

```







```

Db      391 TTAK--MSSSTAENVQLPAEKGVGMDOQVSGEGPHLETAKPTDSAMEQVVEADVAM 447
Oy      55 LPYKGELEKGYQFQDMEISGEGKQDAGYVNLKDTFIKPFKIKBEKKE-----106
Db      448 NPVEKAMSEVVEAGALNPVVEADG-----AMNPVEKAMQGIYEAHMAINQAVD 501
Oy      107 ---BENKPTFVSKKKQNPQ--VNHSQNLNESHKEDLQREHSHQSDSTKDTATVLDKN 161
Db      502 ANPQQAAPTGNDDASDDPSEPVSHS-----ETLNPELEKGVVWRDATERSVSADCCQDN 558
Oy      162 N-----ISSKSTNNPNK 174
Db      559 SKTAAESSLOEISASQTEENSPTQ 582

RESULT 39
ORMSPI
Microtubule-associated protein MAP1B - mouse
N/Alternate names: microtubule-associated protein MAP1(X); microtubule-associated protei
C/Species: Mus musculus (house mouse)
C/Date: 30-Sep-1991 #sequence revision 30-Sep-1991 #text_change 01-Sep-2000
C/Accession: S07549; S44387; A33645
R/Noble: M.; Lewis, S.A.; Cowan, N.J.
J. Cell Biol. 109, 3367-3376, 1989
A/Title: The microtubule binding domain of microtubule-associated protein MAP1B containe
A/Reference number: A33645; MUID:90094539; PMID:2480963
A/Accession: S07549
A/Molecule type: mRNA
A/Residues: 1-2464 <NOB>
A/Cross-references: EMBL:X51396; NID:952999; PIDN:CAA35761.1; PID:953000
R/Sanchez, C.; Padilla, R.; Paciucci, R.; Zabala, J.C.; Avila, J.
Arch. Biochem. Biophys. 310, 428-432, 1994
A/Title: Binding of heat-shock protein 70 (hsp70) to tubulin.
A/Reference number: S44387; MUID:94234720; PMID:8179328
A/Accession: S44387
A/Status: preliminary
A/Molecule type: protein
A/Residues: 653-663, 'IC' <SAN>
C/Superfamily: microtubule-associated protein MAP1B
C/Keywords: microtubule binding; phosphoprotein; tandem repeat
F/589-766/Domain: microtubule binding #status experimental <MTB>
F/589-592, 639-642, 649-652, 655-658, 668-671, 674-677, 679-682, 683-686, 687-690, 691-69
R-K-E/D-X)
F/1861-2064/Region: 17-residue repeats
F/91,116,351,888,1124,1153,1168,1208,1662,1877,1918,2003,2030,2054,2083/Binding site: ph
F/147,969,1336,1562,1563,1702,1708,1990,2057,2066,2419/Binding site: phosphate (Thr) (cc
F/1993/Binding site: phosphate (Tyr) (covalent) #status predicted

Query Match      10.4%; Score 93; DB 1; Length 2464;
Best Local Similarity 27.8%; Pred. No. 1.2e+02;
Matches 49; Conservative 24; Mismatches 57; Indels 46; Gaps 11;

Oy      14 KEFLAKKTGEVSEIKPHRYVTYTONGKMSSTIYSEEPFLPYKGELEKGYQDGMWI 73
Db      584 EKVLVKKQKPVTKESKP--SVT---EKQVSS---KEED--SPV-KAEVA-----BK 623
Oy      74 SGPEKKDAGYVNLKDTFIKPFK--KIEKKKEBENKTPVSKKKQNPQVNHSQLNES 132
Db      624 QATSEKP-----KVTQKQVVKKEIKTKLEKKEE--KPKKVVKKEDTTP--KQDK 672
Oy      133 HKKEDLORE-----EHSQKSDSTKDTATVLDKNNLSSKSTNNPNK 174
Db      673 PKKEVKKKEIKKKEIKKKEIKKKEIKKKEIKKKEIKKKEIKKKEIKKKEIKKKEIKK 728

RESULT 40
Serine-type D-Ala-D-Ala carboxypeptidase (EC 3.4.16.4) daea - Bacillus subtilis
N/Alternate names: penicillin-binding protein 5
C/Species: Bacillus subtilis
C/Date: 28-Oct-1996 #sequence revision 13-Mar-1997 #text_change 20-Jun-2000
C/Accession: S66040; I39830; A922307; B61335; D65612; A23307
R/Ogasawara, N.; Nakai, S.; Yoshikawa, H.

```

```

DNA Res. 1, 1-14, 1994
A/Title: Systematic sequencing of the 180 kilobase region of the Bacillus subtilis chr
A/Reference number: S65967; MUID:96051385; PMID:7584024
A/Accession: S66040
A/Status: preliminary; nucleic acid sequence not shown; translation not shown
A/Molecule type: DNA
A/Residues: 1-443 <OGA>
A/Cross-references: EMBL:D26185; NID:9467326; PIDN:BA05246.1; PID:9467400
A/Note: The nucleotide sequence was submitted to the EMBL Data Library, December 1993
R/Todd, J.A.; Roberts, A.N.; Johnstone, K.; Piggett, P.J.; Winter, G.; Ellar, D.J.
J. Bacteriol. 167, 257-264, 1986
A/Title: Reduced heat resistance of mutant spores after cloning and mutagenesis of the
A/Reference number: I39830; MUID:86250602; PMID:3087956
A/Accession: I39830
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 48-226, 'Q', 228-443 <RES>
A/Cross-references: GB:M13766; NID:9142816; PIDN:AAA22375.1; PID:9142817
R/Waxman, D.J.; Strominger, J.L.
J. Biol. Chem. 255, 3964-3976, 1980
A/Title: Sequence of active site peptides from the penicillin-sensitive D-alanine carb
A/Reference number: A92275; MUID:80182289; PMID:6768745
A/Accession: A92275
A/Molecule type: protein
A/Residues: 32-95, 'X', 97-98, 'XQX', 102 <MAX>
R/Waxman, D.J.; Strominger, J.L.
Proc. Natl. Acad. Sci. U.S.A. 76, 2730-2734, 1979
A/Title: Mechanism of penicillin action: penicillin and substrate bind covalently to cl
A/Reference number: A61335; MUID:79223865; PMID:111240
A/Accession: B61335
A/Molecule type: protein
A/Residues: 55-68 <YOC>
R/Kunert, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berti
C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; C
A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabre, C.; Ferrari, E
Nature 390, 249-256, 1997
A/Authors: Pouliger, D.; Fritze, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Gallizzi, A.; Gall
iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.;
Koeltter, P.; Konigstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapins, A.; Lardinozi
A/Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mau
Y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, B.; Rose, M.; Sadate, Y.; Sato, T.; Scanlon
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, F.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Ser
A/Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Ser
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpestra, P.; Tognoni, A.; Tosato, V.; Uchiyam
T.; Winters, P.; Wipac, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida,
A/Authors: Yoshikawa, H.F.; Zumbstein, B.; Yoshikawa, H.; Danchin, A.
A/Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis
A/Reference number: A69580; MUID:98044033; PMID:9384377
A/Accession: D69612
A/Status: nucleic acid sequence not shown; translation not shown
A/Molecule type: DNA
A/Residues: 1-443 <KUN>
A/Cross-references: GB:Z99104; GB:AL009126; NID:92632267; PIDN:CAB11786.1; PID:9263227
A/Experimental source: strain 168
C/Genetics:
A/Genes: daea
A/Start codon: TTG
C/Superfamily: penicillin-binding protein 5
C/Keywords: cell wall synthesis; hydrolase; serine carboxypeptidase
F/67/Active site: Ser #status experimental

Query Match      10.3%; Score 92.5; DB 2; Length 443;
Best Local Similarity 25.0%; Pred. No. 19;
Matches 36; Conservative 26; Mismatches 49; Indels 33; Gaps 6;

Oy      6 DFA-KNTYKPEFLNDDTGEVSEIKPHRYVTYTONGKMSSTIYSEEPFLPYKGELEK 64

```

Db	304	DYAFDNFSMKRIYAECD-----QVKGHK-TISVDKGEKEVGIVTNKAFSLPVNNGE-EK	356
Qy	65	GYQPDGWEISGFEGKKDAGVIVINLSKDTFKVPFVKIIEKKEBBNKPTFDVSKKKDNPOV	124
Db	357	NYKAK-----VTLNKDNLTAPVKKGTGVGK-----LTAEYTGDEK	391
Qy	125	NHSQLNESHKREDIQREBSQKSD	148
Db	392	DYGFNSDLAGVDLVTKEVNEKAN	415

Search completed: February 10, 2004, 10:58:37  
Job time : 14.5638 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 10, 2004, 10:48:45 ; Search time 11.8417 Seconds

(Without alignments)  
1331.870 Million cell updates/sec

Title: US-10-067-385-8\_COPY\_610\_773

Perfect score: 848  
Sequence: 1 TTVEKFLINKDTGSEVSELPK.....ATVLDKNNISKSTNNPNK 164

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :  
1: PIR1:\*  
2: PIR2:\*  
3: PIR3:\*  
4: PIR4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	848	100.0	2140	2 P95074	serine proteinase,
2	845	99.6	2144	2 A97942	metalloproteinase,
3	111	13.1	1038	2 JC5497	claustrin - chloke
4	110	13.0	558	2 T18467	hypothetical prote
5	106	12.5	665	2 B71609	hypothetical prote
6	104	12.3	540	2 D86432	hypothetical prote
7	103.5	12.2	325	2 T18283	hypothetical prote
8	101.5	12.0	312	2 G81339	probable membrane
9	100	11.8	622	2 A90570	lipoprotein (impor
10	99.5	11.7	385	2 T20410	hypothetical prote
11	99	11.7	211	2 T25911	hypothetical prote
12	99	11.7	219	2 B72291	hypothetical prote
13	97.5	11.5	614	2 A84152	hypothetical prote
14	97.5	11.5	1345	2 S46817	hypothetical prote
15	97	11.4	988	2 T14188	hypothetical prote
16	96.5	11.4	456	2 T05612	hypothetical prote
17	96	11.3	535	2 T37189	hypothetical prote
18	96	11.3	2500	2 G71609	hypothetical prote
19	95.5	11.3	644	2 T47835	hypothetical prote
20	95	11.2	348	2 T37271	cylicin II - human
21	95	11.2	1397	2 T10466	DNA topoisomerase
22	94.5	11.1	867	2 T27136	hypothetical prote
23	94.5	11.1	871	2 T27135	hypothetical prote
24	94.5	11.1	3724	2 T18427	hypothetical prote
25	94	11.1	210	2 T28771	hypothetical prote
26	93.5	11.0	645	2 B89883	conserved hypotet
27	93.5	11.0	649	2 S42488	dnak-type molecula
28	93	11.0	629	2 G96542	hypothetical prote
29	93	11.0	2464	1 QRMSP1	microtubule-associ

30	92.5	10.9	1888	2 T39009	hypothetical prote
31	92	10.8	688	2 A47705	triacylglycerol 11
32	92	10.8	1202	1 S05362	probable DNA-direc
33	92	10.8	1875	2 S38173	myosin-like protei
34	91.5	10.8	650	2 D90093	heat shock protein
35	91	10.7	253	2 T32879	hypothetical prote
36	91	10.7	1087	2 T30330	gelsolin-related p
37	90.5	10.7	443	2 S66040	serine-type D-Ala-
38	90.5	10.7	470	2 T28391	ORF MSV230 hypoch
39	90.5	10.7	1332	2 S41552	probable transcrip
40	90.5	10.7	2401	2 T28676	rhodopy protein -
41	90	10.6	700	2 S67610	probable membrane
42	89.5	10.6	433	2 A89951	trigger factor (im
43	89.5	10.6	487	2 T45982	hypothetical prote
44	89.5	10.6	508	2 B81594	hypothetical prote
45	89.5	10.6	508	2 C72074	hypothetical prote

#### ALIGNMENTS

##### RESULT 1

P95074 serine proteinase, subtilase family [imported] - Streptococcus pneumoniae (strain TIGR4  
C:Species: Streptococcus pneumoniae  
C:Date: 03-Aug-2001 #sequence\_revision 03-Aug-2001 #text\_change 03-Aug-2001

C:Accession: P95074

R:Retcelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Hel  
on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapple,  
nson, T.; Hickey, E.K.; Holt, I.E.  
Science 293, 498-506, 2001

A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison

A:Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.

A:Reference number: A95000; MUID:21357209; PMID:11463916

A:Accession: P95074

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-2140 <KOR>

A:Cross-references: GB:AB005672; PIDN:AAK74791.1; PID:G14972117; GSPDB:GR00164; TIGR:SP

A:Experimental source: strain TIGR4

C:Genetics:

A:Gene: SP0641

Query Match 100.0%; Score 848; DB 2; Length 2140;  
Best Local Similarity 100.0%; Pred. No. 6.5e-54;  
Matches 164; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	TTVEKFLINKDTGSEVSELPKPHRTVTITONGKMSSTVSEEDFLIPVYKGELEKGYQFDG	60
Db	1943	TTVEKFLINKDTGSEVSELPKPHRTVTITONGKMSSTVSEEDFLIPVYKGELEKGYQFDG	2002
Qy	61	WEISGFBRKQDAGVINLSKDTFKPKYKKEKKEBKNNKPTPVSKKKDNPQVHSQLN	120
Db	2003	WEISGFBRKQDAGVINLSKDTFKPKYKKEKKEBKNNKPTPVSKKKDNPQVHSQLN	2062
Qy	121	ESHRKEDLOREHSHQSDSTKQVATVLDKNNISKSTNNPNK	164
Db	2063	ESHRKEDLOREHSHQSDSTKQVATVLDKNNISKSTNNPNK	2106

##### RESULT 2

A97942 metalloproteinase (BC 3.4.21.-) A [imported] - Streptococcus pneumoniae (strain R6)  
C:Species: Streptococcus pneumoniae  
C:Date: 22-Oct-2001 #sequence\_revision 22-Oct-2001 #text\_change 02-Nov-2001

C:Accession: A97942

R:Hoekins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; Dehoff, B.S.;  
R.; Leblanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.;  
Y., P.; Sun, P.M.; Winkler, M.E.  
J. Bacteriol. 183, 5709-5717, 2001

A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaekunas, S.R.

A:Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.

A:Reference number: A97872; MUID:2143245; PMID:11544234

A:Accession: A97942  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-2144 <KUD>  
 A:Cross-references: GB:AE007317; PIDN:AAK9365.1; PID:q15458138; GSPDB:GM00174  
 C:Genetics:  
 A:Gene: pTA  
 C:Keywords: hydrolase; serine proteinase

Query Match 99.6%; Score 845; DB 2; Length 2144;  
 Best Local Similarity 99.4%; Pred. No. 1.1e-53;  
 Matches 163; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTVPKPIKNDGSEVSELPKPRVTYTIQNGKMSSTIVSEEDFILPVYKGELEKGYQFDG 60  
 DB 1947 TTVPKPIKNDGSEVSELPKPRVTYTIQNGKMSSTIVSEEDFILPVYKGELEKGYQFDG 2006  
 QY 61 WEISGPEGKADAGVYNLSKDTPIKPVFKKIBKKKEENKPTDVSKKDNPQVNSQLN 120  
 DB 2007 WEISGPEGKADAGVYNLSKDTPIKPVFKKIBKKKEENKPTDVSKKDNPQVNSQLN 2066  
 QY 121 ESHRKEDLQREHSHQSDSTKDTATVLDKNNISSKSTNNPNK 164  
 DB 2067 ESHRKEDLQREHSHQSDSTKDTATVLDKNNISSKSTNNPNK 2110

## RESULT 3

JCS497

claustrin - chicken

M:Alternate names: keratan sulfate proteoglycan

C:Species: Gallus gallus (chicken)

C:Date: 07-Jul-1997 #sequence, revision 12-Sep-1997 #text\_change 21-Jul-2000

A:Accession: JCS497; PCJ334; S37561

R:Burg, M.A.; Cole, G.J., 1994

J. Neurobiol. 25, 1-22, 1994

A:Title: Claustrin, an antiadhesive neural keratan sulfate proteoglycan, is structurally

A:Reference number: JCS497; PMID:94157526; PMID:7906711

A:Accession: JCS497

A:Molecule type: mRNA

A:Residues: 1-1038 &lt;BUR1&gt;

A:Cross-references: EMBL:X67778; NID:q406318; PIDN:CAA47988.1; PID:q406319

A:Accession: PCJ334

A:Molecule type: protein

A:Residues: 79-83;299-412;485-502 &lt;BUR2&gt;

A:Experimental source: brain

C:Comment: This protein inhibits neural cell adhesion and neurite outgrowth in the nerv

C:Keywords: chondroitin sulfate proteoglycan; glycoprotein; keratan sulfate

F:1267-270/Region: cell attachment (R-G-D) motif

F:112,213,490/Binding site: carbohydrate (Asn) (covalent) #status predicted

F:152,249,440,793,820/Binding site: chondroitin sulfate (Ser) (covalent) #status predict

F:152,249,440,793,820/Binding site: chondroitin sulfate (Ser) (covalent) #status predict

F:152,249,440,793,820/Binding site: chondroitin sulfate (Ser) (covalent) #status predict

F:152,249,440,793,820/Binding site: chondroitin sulfate (Ser) (covalent) #status predict

F:152,249,440,793,820/Binding site: chondroitin sulfate (Ser) (covalent) #status predict

F:152,249,440,793,820/Binding site: chondroitin sulfate (Ser) (covalent) #status predict

F:152,249,440,793,820/Binding site: chondroitin sulfate (Ser) (covalent) #status predict

F:152,249,440,793,820/Binding site: chondroitin sulfate (Ser) (covalent) #status predict

F:152,249,440,793,820/Binding site: chondroitin sulfate (Ser) (covalent) #status predict

F:152,249,440,793,820/Binding site: chondroitin sulfate (Ser) (covalent) #status predict

F:152,249,440,793,820/Binding site: chondroitin sulfate (Ser) (covalent) #status predict

F:152,249,440,793,820/Binding site: chondroitin sulfate (Ser) (covalent) #status predict

F:152,249,440,793,820/Binding site: chondroitin sulfate (Ser) (covalent) #status predict

F:152,249,440,793,820/Binding site: chondroitin sulfate (Ser) (covalent) #status predict

F:152,249,440,793,820/Binding site: chondroitin sulfate (Ser) (covalent) #status predict

F:152,249,440,793,820/Binding site: chondroitin sulfate (Ser) (covalent) #status predict

F:152,249,440,793,820/Binding site: chondroitin sulfate (Ser) (covalent) #status predict

F:152,249,440,793,820/Binding site: chondroitin sulfate (Ser) (covalent) #status predict

F:152,249,440,793,820/Binding site: chondroitin sulfate (Ser) (covalent) #status predict

T18467  
 hypothetical protein C0465c - malaria parasite (Plasmodium falciparum)

C:Species: Plasmodium falciparum

C:Date: 15-Oct-1999 #sequence, revision 15-Oct-1999 #text\_change 11-Jan-2002

C:Accession: T18467

R:Lawson, D.; Bowman, S.; Barrell, B.

submitted to the EMBL Data Library, November 1998

A:Reference number: Z18937

A:Accession: T18467

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-558 &lt;LAW&gt;

A:Cross-references: EMBL:AL008970; NID:e1407852; PIDN:CAA15610.1

A:Map position: 3

A:Introns: 84/1; 160/1

A:Note: C0465c

Query Match 13.0%; Score 110; DB 2; Length 558;  
 Best Local Similarity 29.3%; Pred. No. 1.1;  
 Matches 54; Conservative 23; Mismatches 47; Indels 60; Gaps 12;

QY 37 IYSEEDFILPVY-----KGELEKGYQFDGWEISGPEGK-----DAGVYNLSKDTPIKPV 87  
 DB 60 ILGFEDDILYCYCISQKQSKER-KADGEDRYLNAKKIKNLGTGKKSDFIETL 117  
 QY 88 FKFI--EKKKE-----ENKPTDVS-KKONPVNSQLNE-----SHRK 125  
 DB 118 LELLINERKKEBHIDTLNENK-TNDIKVKNENINENVNENMDISKDKHSHON 176  
 QY 126 E-----DLQREH-----SQKSDSTK---DYATVLDKNNISSKSTTN 160  
 DB 177 EHNINNVNLKKEKTYDIDQDKKHKRSLSQKSDSYKRPFNKRKTSIER-SLSNKKYDE 235  
 QY 161 NPNK 164  
 DB 236 KTKK 239

## RESULT 5

B71609

hypothetical protein PFB0680w - malaria parasite (Plasmodium falciparum)

C:Species: Plasmodium falciparum

C:Date: 13-Nov-1998 #sequence, revision 13-Nov-1998 #text\_change 21-Jul-2000

A:Accession: B71609

R:Gardner, M.J.; Tetteh, H.; Carucci, D.J.; Cummings, L.M.; Ayalind, L.; Koonin, E.V.

; Perle, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.

Science 282, 1126-1132, 1998

A:Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.

A:Reference number: A71600; PMID:99021743; PMID:9804551

A:Accession: B71609

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-665 &lt;GAR&gt;

A:Cross-references: GB:AE001410; GB:AE001362; NID:g3845245; PIDN:AACT1925.1; PID:g3845

A:Experimental source: clone 3D7

A:Gene: PFB0680w

Query Match 12.5%; Score 106; DB 2; Length 665;  
 Best Local Similarity 24.3%; Pred. No. 2.6;  
 Matches 42; Conservative 35; Mismatches 40; Indels 56; Gaps 10;

QY 10 KDTGSEVSELPKPRVT-VTIQNGKMSSTIVSEEDFILPVYKGELEKGYQFDGMI--SGF 66  
 DB 127 EKNKINKSLDHPQNELNLSGK-----NEQDI-----NKNKXKQ-----DISNSNA 169  
 QY 67 EKKKADAGVYNLSKDTPIKPVFKKIBKKK-----EENKPTD-----VSKKONP 112  
 DB 170 ENKKD-----VKGAVKELERKKKBEKISDHYKVENKSDHYKVENKSDDH 217  
 QY 113 QVNSQLNSHSHRKEDLQREHSHQSDSTKDTATVLDKNNISSKSTNNPNK 164

## RESULT 4

```
Query Match      12.2% Score 103.5; DB 2; Length 325;  
Best Local Similarity 23.8%, Pred. No. 1.7;  
Matches 43; Conservative 25; Mismatches 50; Indels 63; Gaps 9.  
  
Oy   21 HRTVTTINGKEMSETIVSEBDPIIPYVK-GL--EKGYDPGMELSGFECK----- 69  
    || :|{| |::||::|||::|||::|||  
Db   57 HRTTISLN--RFSVKIKIGDEKLRIKSNQELIYLNLLEFPNFIATK--EGKHILAKSNWF 112  
    ||::||::||::||::||::||:  
  
Oy   70 --KDAGTV-----INLSKDTFKPV-----PKX----- 90
```

A, Genetic code: SGC3

Query Match 11.84; Score 100; DB 2; Length 622;

A, Introns: 30/2; 200/3

Best Local Similarity 27.3%; Pred. No. 9.8;









A: Molecule type: DNA  
A: Residues: 1-649 <RES>  
A: CROSB: references: EMBL:X72621, NID:g461335, PIDN:CAA5197.1, PID:g461336  
R: Hochmann, C.U.B.; Remington, S.A.; Haebler, M.M.; Martin, W.F.; Mueller, S.B.; Couch, J.  
Mol. Gen. Genet. 243, 600-604, 1994  
A: Title: The smallest known eukaryotic genomes encode a protein gene: towards an understan  
A: Reference number: S45576; NUID:94268506; PMID:8208251

Query Match	11.0%;	Score 93;	DB 2;	Length 629;
Best Local Similarity	19.6%;	Pred. No. 21;		
Matches	40;	Conservative	39;	Mismatches 73; Indels 52; Gaps 7;
Qy	1	TTVKEFLINKQGEV-----	SELKPHRYVTYIIONGKENSSTIVSEDFI	44
		:::	:::	
Db	391	TTAK---MSSSTAFQQLPAKGVGKMDQKVSQGMHLEFAKTKDSAMQVTEABDYAM		447
Qy	45	LPVYKGELEKGYQPDGWEISGFEKKDAGVYINLSKDTYIKPFPKYLEKCKE-----		96
		:::	:::	
Db	448	NPIVEKAMSEWYEAEGAIINPIVEAEDG-----ANNPIVEKAMSOIVEAEDAINDAVD		501
Qy	97	---EENKPTPDVSKKKDNPQ--VNNSQLNESHKRELQREBSQKSDSTQDTATVATVDKN		151
		:::	:::	
Db	502	ANFQALPQAGNDADSDSDPSBPVSHS--ETLNPPLEKKEVWRKQATERSVSLDCCDN		556

Qy 152 N-----ISSKSTNNPK 164  
 Db 559 SKTIAESSLQEIETISASQTEHSPTQ 582

## RESULT 29

ORMSPI

microtubule-associated protein MAP1B - mouse  
 N/Alternate names: microtubule-associated protein MAP1(X); microtubule-associated protei

C/Species: Mus musculus (house mouse)  
 C/Date: 30-Sep-1991 #sequence revision 30-Sep-1991 #text\_change 01-Sep-2000  
 C/Accession: S07549; S44387; A33645  
 R/Mobile, M.; Lewis, S.A.; Cowan, N.J.  
 J. Cell Biol. 109, 3367-3376, 1989

A/Title: The microtubule binding domain of microtubule-associated protein MAP1B contains  
 A/Reference number: A33645; MUID:90094559; PMID:2480963  
 A/Accession: S07549

A/Molecule type: mRNA  
 A/Residues: 1-2464 <NOB>  
 A/Cross-references: EMBL:X51396; NID:952999; PIDN:CAA35761.1; PID:953000  
 R/Sanchez, C.; Padilla, R.; Paciucci, R.; Zabala, J.C.; Avila, J.  
 Arch. Biochem. Biophys. 310, 428-432, 1994  
 A/Title: Binding of heat-shock protein 70 (hsp70) to tubulin.  
 A/Reference number: S44387; MUID:94234720; PMID:8179328  
 A/Accession: S44387

A/Status: preliminary  
 A/Molecule type: protein  
 A/Residues: 653-663; 'IC' <SAN>  
 C/Superfamily: microtubule binding; phosphoprotein; tandem repeat  
 C/Keywords: microtubule binding; phosphoprotein; status experimental <MTB>  
 F:589-592,639-642,649-652,655-658,660-663,668-671,674-677,679-682,683-686,687-690,691-69  
 R-K-E/D-X)  
 F:1861-2064/Region: 17-residue repeats  
 F:91,116,351,888,1124,1153,1168,1208,1662,1877,1918,2003,2030,2054,2083/Binding site: pH  
 F:147,969,1336,1562,1553,1702,1708,1990,2057,2063,2419/Binding site: phosphate (Thr) (cc  
 F:11953/Binding site: phosphate (Tyr) (covalent) #status predicted

Query Match 11.0%; Score 93; DB 1; Length 2464;  
 Best Local Similarity 27.8%; Pred. No. 1e+02;  
 Matches 49; Conservative 24; Mismatches 57; Indels 46; Gaps 11;

Qy 4 KEPIIKNDGTGEVSELKPHRVTTIIONGKMSSTIVSEEDFILPVYKGELEKGYQPDGWEI 63  
 Db 584 EKVLVAKDKRPVKTESKPD---SVT---EKEVSS---KEEQ---SPV-KAEVA-----EK 623  
 Qy 64 SGEFGKDGAGYVINSKDTFIKPVFK-KIEKKEENKPTFDVSKKKNPQVNSQLNES 122  
 Db 624 QATESKPD-----KVTKDKVVKKEIKTKLEKKEK--KPKCGVVKCKEDKTPPL---KKDEK 672  
 Qy 123 HKREDIQR-----EHSQKSDSTKDVATVLDKNNISSKSTNNPK 164  
 Db 673 PRKEVYKEIKKEIKKEERKELKKEVKEPTPLDAAKKEVKEKKEKKEKPK 728

## RESULT 30

T39009

hypothetical protein SPAC6B12.02c - fission yeast (Schizosaccharomyces pombe)  
 C/Species: Schizosaccharomyces pombe

C/Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 21-Jan-2000  
 C/Accession: T39009  
 R/Gentile, S.; Churcher, C.M.; Barrell, B.G.; Rajindream, M.A.; Wood, V.  
 submitted to the EMBL Data Library, August 1995  
 A/Reference number: 221815

A/Accession: T39009  
 A/Status: preliminary; translated from GB/EMBL/DDDBJ  
 A/Molecule type: DNA

A/Residues: 1-1888 <GEN>  
 A/Cross-references: EMBL:Z98531; PIDN:CAB11064.1; GSPDB:GN00066; SPDB:SPAC6B12.02c

C/Genetics:  
 A/Genetic source: strain 972h-; cosmid c6B12  
 A/Map position: 1

C/Superfamily: Schizosaccharomyces hypothetical protein SPAC6B12.02c

Query Match 10.9%; Score 92.5; DB 2; Length 1888;  
 Best Local Similarity 23.1%; Pred. No. 81;  
 Matches 39; Conservative 25; Mismatches 76; Indels 29; Gaps 5;

Qy 1 TTVEPIIKNDGTGEVSELKPHRVTTIIONGKMSSTIVSE-BDFILPVYKGELEKGYQPD 59  
 Db 394 TSENPFQLVANVANNASTIVRTTKTKKGNRPKYVEVELPLILLESY----- 442  
 Qy 60 GWEISGFEKDGAGYVINSKDTFIKPVFKIEKKEENKPTFDVSKKKNPQVNSQL 119  
 Db 443 -----GKAPKPLRVFARSSSHIP--KMLPRKQKMSKKYFSPDKSDRQVLDVLS 492  
 Qy 120 NESHREDIQRBSHSQKSDS-TKDVATVLDKN-----NISKSTNN 161  
 Db 493 DWYSGKHELIVQOSHYSKPKSDSKSVGNIFSVNSKKSIVINAKTANN 541

## RESULT 31

A47705

triacylglycerol lipase (EC 3.1.1.3) - Staphylococcus epidermidis

C/Species: Staphylococcus epidermidis  
 C/Date: 19-Dec-1993 #sequence\_revision 18-Nov-1994 #text\_change 08-Oct-1999  
 C/Accession: A47705

R/Farrell, A.M.; Foster, T.J.; Holland, K.T.  
 J. Gen. Microbiol. 139, 267-277, 1993

A/Title: Molecular analysis and expression of the lipase of Staphylococcus epidermidis.  
 A/Reference number: A47705; MUID:93171870; PMID:8436947

A/Content: 9  
 A/Accession: A47705

A/Status: preliminary  
 A/Molecule type: protein

A/Residues: 1-688 <PAR>  
 A/Cross-references: GB:M95577; NID:g153021; PIDN:AAA19729.1; PID:g153022

A/Note: sequence extracted from NCBI backbone (NCBI:125632, NCBI:P.125633)  
 C/Superfamily: Staphylococcus triacylglycerol lipase

C/Keywords: carboxylic ester hydrolase

Query Match 10.8%; Score 92; DB 2; Length 688;  
 Best Local Similarity 25.0%; Pred. No. 28;  
 Matches 40; Conservative 28; Mismatches 66; Indels 26; Gaps 7;

Qy 9 NMDTGEVSELKPHRVTTIIONGKMSSTIVSEEDFILPVYKGELEKGYQPDGWEI 68  
 Db 68 NKNVNEKSNVN-----SITNESLHNPTKNDLI-----QQQKDSQNDNSSESVEQ 115  
 Qy 69 KQDAG-YVINSKDTFIKPVFKIE-EKKEENKPTFDVSKKKNPQV--NHSQLAESHR 124  
 Db 116 NKENGAFFVQNHSEB--KQDQGVLEIKHASNNQTLHSKAAQSNEDVTKTKPSQLDWTAA 172  
 Qy 125 KEDLQRBSHSQKSDSTKDVATVLDKNNISSKSTNNPK 164  
 Db 173 KQEDSQKEMLSKQDTOSSTKTDLL-----RATAQNSK 205

## RESULT 32

S05362

probable DNA-directed DNA polymerase (EC 2.7.7.7) - fungus (Ascobolus immersus) mitochor

C/Species: mitochondrion Ascobolus immersus  
 C/Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 16-Jun-2000

C/Accession: S05362  
 R/Kempken, F.; Meinhardt, F.; Esser, K.  
 Mol. Gen. Genet. 218, 523-530, 1989

A/Title: In organello replication and viral affinity of linear, extrachromosomal DNA of

A/Reference number: S05362; MUID:9006356; PMID:2573821

A/Accession: S05362  
 A/Molecule type: DNA

A/Residues: 1-1202 <KEM>  
 A/Cross-references: EMBL:X15982; NID:g2933; PIDN:CAA4106.1; PID:g1370212

C/Genetics:  
 A/Genetic source: mitochondrion  
 A/Genetic code: SGC3







```
Db      924  TSXNHEEKISDIRKNSIKITQDFSEBSYINDI-KKELEKNVLESQNNNTDINOYLSKIBN 982
Qy      73   GYVINLSKOTPIKPVFKKIEBKEENKPTPDVSKKQDNQVNHQSQLNESHKEDLOREE 132
Db      983  IY--NILKLNKKIKKIIDKVKYTDIEK----NNKKINAEIJSNKKIITQLEKNSLIKE 1035
Qy      133  HSQKSDSTKD-----VTATVLDKNNISS 155
Db      1036  QSKKIKSTIIDNVYSECINKITNLKTYIVNEKNNINT 1072
```

Search completed: February 10, 2004, 10:58:35  
Job time : 12.8417 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 10, 2004, 10:48:44 ; Search time 29.0366 Seconds  
(without alignments)  
1457.493 Million cell updates/sec

Title: US-10-067-385-8\_COPY\_610\_773

Perfect score: 848  
Sequence: 1 TTYKRPILNDTGEVSEIKP.....ATVLDKNISKSTTNPNK 164

Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

SPTREMBL\_23:\*

- 1: sp\_archaea:\*
- 2: sp\_bacteria:\*
- 3: sp\_fungi:\*
- 4: sp\_human:\*
- 5: sp\_invertebrate:\*
- 6: sp\_mammal:\*
- 7: sp\_mhc:\*
- 8: sp\_organelle:\*
- 9: sp\_phage:\*
- 10: sp\_plant:\*
- 11: sp\_rodent:\*
- 12: sp\_virus:\*
- 13: sp\_vertebrate:\*
- 14: sp\_unclassified:\*
- 15: sp\_virus:\*
- 16: sp\_bacteriopl:\*
- 17: sp\_archaeopl:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	848	100.0	2119	2	Q9AHT5 streptococc
2	848	100.0	2140	16	Q97RY6 streptococc
3	845	99.6	2144	16	Q8dnp7 streptococc
4	844	99.5	2144	2	Q9s4m8 streptococc
5	119	14.0	775	16	Q8cpk8 staphylococ
6	112.5	13.3	361	5	Q95p15 plasmodium
7	112.5	13.3	379	5	Q9u6c4 plasmodium
8	111.5	13.1	346	5	Q9u0g0 plasmodium
9	111.5	13.1	354	5	Q25995 plasmodium
10	111.5	13.1	354	5	Q81J55 plasmodium
11	111.5	13.1	379	5	Q25706 plasmodium
12	111	13.1	1038	13	Q90784 gallus galli
13	110.5	13.0	379	5	Q25705 plasmodium
14	110	13.0	829	5	Q815f3 plasmodium
15	109.5	12.9	609	5	Q812K8 plasmodium
16	109	12.9	380	5	Q26019 plasmodium

17	109	12.9	3008	5	Q81436 plasmodium
18	108.5	12.8	600	5	Q77355 plasmodium
19	107.5	12.7	470	10	Q9FVK9 arabidopsis
20	106	12.5	951	5	Q96229 plasmodium
21	104.5	12.3	329	5	Q9NPF9 arabidopsis
22	104	12.3	540	10	Q94CS9 arabidopsis
23	104	12.3	540	10	Q9SA84 arabidopsis
24	103.5	12.2	325	5	Q44016 dictyosteli
25	103	12.1	2081	10	Q9LH98 arabidopsis
26	102.5	12.1	238	5	Q81226 plasmodium
27	101.5	12.0	312	16	Q9PPI5 campylobact
28	101.5	12.0	382	5	Q9V7J0 drosophila
29	101.5	12.0	556	5	Q9V7I9 drosophila
30	101.5	12.0	556	5	Q9G082 drosophila
31	101.5	12.0	948	5	Q9U4U6 plasmodium
32	101.5	12.0	1377	5	Q81566 plasmodium
33	100.5	11.9	1373	5	Q81353 plasmodium
34	100	11.8	375	4	Q14712 homo sapien
35	100	11.8	622	16	Q98QX1 mycoplasma
36	100	11.8	1130	5	Q81J24 plasmodium
37	100	11.8	3127	5	Q81DA0 caenorhabdi
38	99.5	11.7	385	5	Q9J424 plasmodium
39	99.5	11.7	4524	5	Q813J9 caenorhabdi
40	99	11.7	211	5	P91488 thermotoga
41	99	11.7	219	16	Q9X0W6 thermotoga
42	99	11.7	1859	5	Q81C27 plasmodium
43	99	11.7	2162	5	Q81BH2 plasmodium
44	99	11.7	2563	5	Q813A0 plasmodium
45	99	11.7	3026	5	Q81LS9 plasmodium

#### ALIGNMENTS

RESULT 1

Q9AHT5 PRELIMINARY: PRT; 2119 AA.

AC Q9AHT5; 01-JUN-2001 (TRENBLREL. 17, Created)

DT 01-JUN-2001 (TRENBLREL. 17, Last sequence update)

DT 01-MAR-2003 (TRENBLREL. 23, Last annotation update)

DE Serine protease (Fragment).

GN PR1A.

OS Streptococcus pneumoniae.

OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;

OC Streptococcus.

OX NCBI\_TaxID=1313;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=N4;

RX MEDLINE=21116976; PubMed=11179332;

RA Witzmann T.M., Heinrichs J.H., Adamou J.B., Erwin A.L., Kunsch C., Choi G.H., Barash S.C., Rosen C.A., Masuie H.R., Tuomenen E., Langermann S., Johnson S., Koenig S., Gayle A., Breach Y.A., Walsh W., Barren P., Lachigra R., Hanson M., Langermann S., Johnson S., Koenig S., "Use of a Whole Genome Approach to Identify Vaccine Molecules Affording Protection Against Streptococcus pneumoniae Infection.", Infect. Immun. 69:1593-1596(2001).

RL Affording Protection Against Streptococcus pneumoniae Infection.", -I- SUBCELLULAR LOCATION: ATTACHED TO THE CELL WALL. PEPTIDOGLYCAN BY AN AMIDE BOND (BY SIMILARITY).

CC EMBL: AF291699; AAK19159.1; --

CC HSRP; P00782; 2SBR.

DR MEROPS: S08.064; --

DR InterPro: IPR001899; Gram\_pos\_anchor.

DR InterPro: IPR006192; LPXTG.

DR InterPro: IPR003137; PA.

DR InterPro: IPR002029; Peptidase\_S8.

DR InterPro: IPR001680; WD40.

DR Pfam: PF00746; Gram\_pos\_anchor; 1.

DR Pfam: PF02225; PA; 1.

DR Pfam: PF00082; Peptidase\_S8; 2.

DR PRINTS: PR00723; SUBTILISIN.

DR TIGRPFAMs: TIGR01167; LPXTG\_anchor; 1.



DR PROSITE; PS50847; GRAM\_POS\_ANCHORING; 1.  
DR PROSITE; PS50840; PA; 1.  
DR PROSITE; PS00137; SUBTLASE\_HIS; 1.  
DR PROSITE; PS00138; SUBTLASE\_SER; 1.  
DR PROSITE; PS00678; WD\_REPEATS\_1; 1.  
KM Cell wall, peptidoglycan-anchor; Protease.  
FT NON\_TER 1 1  
SQ SEQUENCE 2119 AA; 238226 MW; 517987696806Aa CRC64;  
  
Query Match 100.0%; Score 848; DB 2; Length 2119;  
Best Local Similarity 100.0%; Pred. No. 9,6e-53;  
Matches 164; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 TTVKEFLINKDTGSEVSELKPHRVTVTTIQNGKMSSTIVSEEDFILPVYKGELEKGYQDFG 60  
DB 1922 TTVKEFLINKDTGSEVSELKPHRVTVTTIQNGKMSSTIVSEEDFILPVYKGELEKGYQDFG 1981  
QY 61 WEISGFEKGKADGYVINLSKDTFIPVKFKIEEKEENKPPFDVSKKKDNPOVNHSSQLN 120  
DB 1982 WEISGFEKGKADGYVINLSKDTFIPVKFKIEEKEENKPPFDVSKKKDNPOVNHSSQLN 2041  
QY 121 ESHRKEDLQREBHSQKSDSTKDVATVLDKNNISSKSTNNPNK 164  
DB 2042 ESHRKEDLQREBHSQKSDSTKDVATVLDKNNISSKSTNNPNK 2085  
  
RESULT 2  
Q97RY6 PRELIMINARY; PRT; 2140 AA.  
ID Q97RY6;  
AC Q97RY6;  
DT 01-OCT-2001 (TrEMBLrel. 18, Created)  
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE Serine protease, subtilase family.  
GN SPO641.  
OS Streptococcus pneumoniae.  
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=1313;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX STRAIN=TRIG4;  
MEDLINE=21357209; PubMed=11463916;  
RA Terelein H., Nelson K.E., Paulsen I.T., Bisen J.A., Read T.D.,  
RA Peterson S., Heidelberg J., Deboy R.T., Haft D.H., Dodson R.D.,  
RA Durkin A.S., Gwin M., Kolonay J.F., Nelson M.C., Peterson J.D.,  
RA Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,  
RA McDonauld L.A., Khouri H., Wolf A.M., Usterback T.R., Hansen C.L.,  
RA Holt I.B., Loftus B.J., Yang F., Smith H.O., Venter J.C.,  
RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;  
RA "Complete genome sequence of a virulent isolate of Streptococcus  
RT pneumoniae";  
RL Science 293:498-506(2001).  
DR EMBL; AB007373; AAK74791.1; -.  
DR MEROPS; S08.064; -.  
DR TIGR; SP0641; -.  
DR InterPro; IPR001899; Gram\_pos\_anchor.  
DR InterPro; IPR006192; LPXTG.  
DR InterPro; IPR003137; PA.  
DR InterPro; IPR000209; Peptidase\_S8.  
DR InterPro; IPR001680; WD40.  
DR Pfam; PF00746; Gram\_pos\_anchor; 1.  
DR Pfam; PF02225; PA; 1.  
DR Pfam; PF00082; Peptidase\_S8; 2.  
DR PRINTS; PR00723; SUBTLISIN  
DR TIGRPFAM; TIGR01167; LPXTG\_anchor; 1.  
DR PROSITE; PS50847; GRAM\_POS\_ANCHORING; 1.  
DR PROSITE; PS50840; PA; 1.  
DR PROSITE; PS00137; SUBTLASE\_HIS; 1.  
DR PROSITE; PS00138; SUBTLASE\_SER; 1.  
DR PROSITE; PS00678; WD\_REPEATS\_1; 1.  
KW Protease; Complete proteome.

SQ SEQUENCE 2140 AA; 240426 MW; FA44ADB2938B334 CRC64;  
  
Query Match 100.0%; Score 848; DB 16; Length 2140;  
Best Local Similarity 100.0%; Pred. No. 9,7e-53;  
Matches 164; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 TTVKEFLINKDTGSEVSELKPHRVTVTTIQNGKMSSTIVSEEDFILPVYKGELEKGYQDFG 60  
DB 1943 TTVKEFLINKDTGSEVSELKPHRVTVTTIQNGKMSSTIVSEEDFILPVYKGELEKGYQDFG 2002  
QY 61 WEISGFEKGKADGYVINLSKDTFIPVKFKIEEKEENKPPFDVSKKKDNPOVNHSSQLN 120  
DB 2003 WEISGFEKGKADGYVINLSKDTFIPVKFKIEEKEENKPPFDVSKKKDNPOVNHSSQLN 2062  
QY 121 ESHRKEDLQREBHSQKSDSTKDVATVLDKNNISSKSTNNPNK 164  
DB 2063 ESHRKEDLQREBHSQKSDSTKDVATVLDKNNISSKSTNNPNK 2106  
  
RESULT 3  
Q8DQ7 PRELIMINARY; PRT; 2144 AA.  
ID Q8DQ7;  
AC Q8DQ7;  
DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE Cell wall-associated serine protease Pta (EC 3.4.21.-).  
GN PRTA OR SPR0561.  
OS Streptococcus pneumoniae (strain ATCC BAA-255 / R6).  
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=171101;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=21429245; PubMed=11544234;  
RA Hoskins J., Albom W.E., Jr., Arnold J., Blaszcak L.C., Burgett S.,  
RA Dehoff B.S., Estrom S.T., Fritz L., Fu D.-J., Fuller W., Geringer C.,  
RA Gilmore R., Glass J.S., Khoja H., Kraft A.R., Lagace R.E.,  
RA LeBlanc D.J., Lee L.N., Lefkowitz E.J., Lu J., Matsushima P.,  
RA McAnen S.M., McHenry M., McLeaster K., Mundy C.W., Nicas T.I.,  
RA Norris F.H., O'Gara M., Peery R.B., Robertson G.T., Rockey P.,  
RA Sun P.-M., Winkler M.E., Yang Y., Young-Bellido M., Zhao G.,  
RA Zook C.A., Baltz R.H., Jaskunas S.R., Roestek P.R., Jr., Skatrud P.L.,  
RA Glas J.I.;  
RA "Genome of the bacterium Streptococcus pneumoniae strain R6";  
RL J. Bacteriol. 183:5709-5717(2001).  
DR EMBL; AE008434; AAK9365.1; -.  
KW Hydrolyase; Complete proteome.  
SQ SEQUENCE 2144 AA; 240436 MW; 8C1B4B1DBC503A0C CRC64;  
  
Query Match 99.4%; Score 845; DB 16; Length 2144;  
Best Local Similarity 99.4%; Pred. No. 1,6e-52;  
Matches 163; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 TTVKEFLINKDTGSEVSELKPHRVTVTTIQNGKMSSTIVSEEDFILPVYKGELEKGYQDFG 60  
DB 1947 TTVKEFLINKDTGSEVSELKPHRVTVTTIQNGKMSSTIVSEEDFILPVYKGELEKGYQDFG 2006  
QY 61 WEISGFEKGKADGYVINLSKDTFIPVKFKIEEKEENKPPFDVSKKKDNPOVNHSSQLN 120  
DB 2007 WEISGFEKGKADGYVINLSKDTFIPVKFKIEEKEENKPPFDVSKKKDNPOVNHSSQLN 2066  
QY 121 ESHRKEDLQREBHSQKSDSTKDVATVLDKNNISSKSTNNPNK 164  
DB 2067 ESHRKEDLQREBHSQKSDSTKDVATVLDKNNISSKSTNNPNK 2110  
  
RESULT 4  
Q9S4M8 PRELIMINARY; PRT; 2144 AA.  
ID Q9S4M8;  
AC Q9S4M8;  
DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)

DT 01-MAR-2003 (Tremblrel. 23, last annotation update)  
 DE Cell wall-associated serine proteinase PrtA precursor.  
 GN PrtA.  
 OS Streptococcus pneumoniae.  
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 OC Streptococcus.  
 OX NCBI\_TaxID=1313;  
 RN [1]  
 RP SOURCE FROM N.A.  
 RC STRAIN=3.B;  
 RA Bette G., Ten Thoren B., Bongaerts R.J.M., Heinz H.-P., Ziek G.;  
 RT "Cloning and sequencing of a novel surface protease of Streptococcus pneumoniae";  
 RT Submitted (Aug-2000) to the EMBL/GenBank/DBJ databases.  
 RL SUBCELLULAR LOCATION: ATTACHED TO THE CELL WALL PEPTIDOLYCAN BY  
 CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE CELL WALL PEPTIDOLYCAN BY  
 CC AN AMIDE BOND (BY SIMILARITY).  
 CC EMBL: AF127143; AAD48399.1; -;  
 CC HSSP: P00782; 2SBT.  
 DR MEROPS: S08\_064; -;  
 DR InterPro: IPR001899; Gram\_pos\_anchor.  
 DR InterPro: IPR006192; LPXTG.  
 DR InterPro: IPR003137; PA.  
 DR InterPro: IPR000209; Peptidase\_S8.  
 DR InterPro: IPR001680; WD40.  
 DR Pfam: PF00746; Gram\_pos\_anchor; 1.  
 DR Pfam: PF02225; PA; 1.  
 DR Pfam: PF00082; Peptidase\_S8; 2.  
 DR PRINTS: PRO0723; SUBTILISIN.  
 DR TIGRfam: TIGR01167; LPXTG\_anchor; 1.  
 DR PROSITE: PS0847; GRAM\_POS\_ANCHORING; 1.  
 DR PROSITE: PS0840; PA; 1.  
 DR PROSITE: PS00137; SUBTILASE\_HIS; 1.  
 DR PROSITE: PS00138; SUBTILASE\_SER; 1.  
 DR PROSITE: PS00678; WD\_REPEATS\_1; 1.  
 DR PROSITE: PS00678; WD\_REPEATS\_1; 1.  
 KW Cell wall; peptidoglycan-anchor; Signal.  
 FT SIGNAL 1 19 POTENTIAL.  
 FT CHAIN 20 2144 CELL WALL-ASSOCIATED SERINE PROTEINASE  
 FT PRTH.  
 SQ SEQUENCE 2144 AA; 240724 MW; 2052511470741331 CRC64;  
 Query Match 99.5%; Score 844; DB 2; Length 2144;  
 Best Local Similarity 98.8%; Pred. No. 1.9e-52;  
 Matches 162; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 TTYKEFLNKDGEVSELKPHRYVTIIONGKMSSTIVSEDFILPYKGELEKGYQFDG 60  
 DB 1947 TTYKEFLNKDGEVSELKPHRYVTIIONGKMSSTIVSEDFILPYKGELEKGYQFDG 2006  
 QY 61 WEISGREGKKDAGVIMLSKDTFIKPVFKKIEKKKEENKPTFDVSKKKONPVYNSQLN 120  
 DB 2007 WEISGREGKKDAGVIMLSKDTFIKPVFKKIEKKKEENKPTFDVSKKKONPVYNSQLN 2066  
 QY 121 BSHRKEDLOREHSQKSDSTKDVATVLDKNNISSKSTNNPNK 164  
 DB 2067 BSHRKEDLOREHSQKSDSTKDVATVLDKNNISSKSTNNPNK 2110  
 RESULT 5  
 Q8CPK8 PRELIMINARY; PRT; 775 AA.  
 ID O8CPK8;  
 AC O8CPK8;  
 DT 01-MAR-2003 (Tremblrel. 23, Created)  
 DT 01-MAR-2003 (Tremblrel. 23, last sequence update)  
 DT 01-MAR-2003 (Tremblrel. 23, last annotation update)  
 DE Penicillin-binding protein 1.  
 GN SR0856.  
 OS Staphylococcus epidermidis.  
 OC Bacteria; Firmicutes; Bacillales; Staphylococcus.  
 OX NCBI\_TaxID=1282;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 12228;  
 RA Zhang Y., Ren S., Li H., Fu G., Lu L., Lu G., Jia J., Tu Y., Qin Z.,

RA Chen Z., Wen Y.;  
 RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.  
 RN EMBL: AB016746; AAO04453.1; -;  
 KM Complete proteome.  
 SQ SEQUENCE 775 AA; 86354 MW; 4A00563A7BB8777C CRC64;  
 Query Match 14.0%; Score 119; DB 16; Length 775;  
 Best Local Similarity 27.0%; Pred. No. 1.1;  
 Matches 47; Conservative 24; Mismatches 57; Indels 46; Gaps 9;  
 QY 10 KDTGEVSELKPHRYVTIIONGKMSSTIVSEDFILPYK-----GELEKGYPDGM-- 61  
 DB 614 EDSVNAQSLKPF-----ITIGKQIKQSVSGTKVLPKSHVMMLTDELTP--DMTGPTK 668  
 QY 62 -EISGFE-----GKKDAGVIM--LSKDTFIKPVFKKIEKKKEENKPTFDV-----K 107  
 DB 669 EDVLAPEDLTKIKVSTKGMFPVNTQSIKQIITK-----NKKIEVLSAED 715  
 QY 108 KKNPQVNSQLNBSHRKEDLOREHSQKSDSTKDVATVLDKNNISSKSTNN 161  
 DB 716 TDDQKTDDESDSNKSKDKADHDHNSSTNN-----DKSNADSKDSDSD 763  
 RESULT 6  
 Q95PT5 PRELIMINARY; PRT; 361 AA.  
 ID Q95PT5;  
 AC Q95PT5;  
 DT 01-DEC-2001 (Tremblrel. 19, Created)  
 DT 01-DEC-2001 (Tremblrel. 19, last sequence update)  
 DT 01-OCT-2002 (Tremblrel. 22, last annotation update)  
 DE Merozoite surface protein 3 (fragment).  
 OS Plasmodium falciparum.  
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
 OX NCBI\_TaxID=5833;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=FVO;  
 RA Hissada H., Saul A., Long C.A., Miller L.H., Stowers A.W.;  
 RT "Merozoite Surface Protein 3 and Protection Against Malaria in Aotus monkeys";  
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AY044180; AAK94780.1; -;  
 FT NON\_TER 1 1  
 FT NON\_TER 361 361  
 SQ SEQUENCE 361 AA; 41163 MW; 6127A3041587BA74 CRC64;  
 Query Match 13.3%; Score 112.5; DB 5; Length 361;  
 Best Local Similarity 22.4%; Pred. No. 1.4;  
 Matches 43; Conservative 33; Mismatches 55; Indels 61; Gaps 8;  
 QY 11 DTGEVSELKPHRYVTIIONGKMSSTIVSEDFILPYKGELEKGYQ----- 57  
 DB 147 ETGE-----RNSRNNFYTKYE-----YAKGVEDYERAKAYQKAAQAV 187  
 QY 58 -----FD-----GWEISGF--EGKKDAG-----YVNLKDTFIKPVFKKIEKKKEEN 99  
 DB 188 LKAKAASVYIILGMRGGGVPRHKKENMLSHLYSSKDKENISKENDVDLDS--KEBEA 246  
 QY 100 KPTFDVSKKKONPVYNSQLNBSHRKEDLOREHSQKSDSTKDVATVLDN----- 151  
 DB 247 EETEEBELKKEENKBEETSEISEDEEBEKEBEENDKKGEQEKSONENNDOCKOME 306  
 QY 152 --NISKSTNN 161  
 DB 307 AONLISHKONNN 318  
 RESULT 7  
 Q9U6C4 PRELIMINARY; PRT; 379 AA.  
 ID Q9U6C4;  
 AC Q9U6C4;  
 DT 01-MAY-2000 (Tremblrel. 13, Created)  
 DT 01-MAY-2000 (Tremblrel. 13, last sequence update)

```

DT 01-OCT-2002 (TReMBLrel. 22, last annotation update)
DB Polymorphic antigen.
GN MSP-3.
OS Plasmidium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmidium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FCCL/NH;
RA Li X.R., Yu X.B., Luo S.H., Shan Z.X., Fang J.M., Xu J.;
RT "Sequence of Plasmidium falciparum secreted polymorphic antigen
  gene".
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF188190; AAF04099.1; -.
SQ SEQUENCE 379 AA; 43316 MW; C152A54E1F9D5F25 CRC64;

Query Match 13.3%; Score 112.5; DB 5; Length 379;
Best Local Similarity 23.6%; Pred. No. 1.5; Mismatches 59; Indels 39; Gaps 8;
Matches 41; Conservative 35;

DY 19 KPHRYVTITONGKEMSTIVSEDF-----ILPYKGLKGYQFD-GWEISGF- 66
DB 171 KPSRINLPSRKTKYAEQV--EKDYERAKNAVQKANOVLKAKKASSYVILGMEFGGV 228
OY 67 -EGKQDAG-----YVNLSDKTFIKPVFKIEKKKEENKPTDVSKKKDPQVNHSQLN 120
DB 229 PEHKKEEMNLSTLYSSKDKENISKENDVDLDE-KEESELEFTEBELEKNEETSEIS 287
OY 121 -----ESHRKEDLQREHSHOKSDSTQVATVATLDKNNISKSTNN 161
DB 288 EDEEEEBEKEKEENEKKEKEQKQENENNDQKQMEA-----QNLISKQNNN 336

RESULT 8
O9U0G0 PRELIMINARY; PRT; 346 AA.
AC Q9U0G0.
DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-MAY-2000 (TReMBLrel. 13, last sequence update)
DT 01-OCT-2002 (TReMBLrel. 22, last annotation update)
DB Merzoite surface protein 3 (Fragment).
GN MSP3.
OS Plasmidium reichenowi.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmidium.
OX NCBI_TaxID=5854;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=20416497; PubMed=10960178;
RA Okumu D.M.N., Thomas A.W., Conway D.J.;
RT "Allelic lineages of the merozoite surface protein 3 gene in
  Plasmidium reichenowi and plasmidium falciparum.";
RL Møl. Biochem. Parasitol. 109:185-188 (2000).
DR EMBL; AJ252286; CAB65754.1; -.
FT NON_TER 1
FT TER 1
SQ SEQUENCE 346 AA; 39127 MW; AB04B96BDFAPAO10 CRC64;

Query Match 13.1%; Score 111.5; DB 5; Length 346;
Best Local Similarity 23.9%; Pred. No. 1.6;
Matches 42; Conservative 32; Mismatches 67; Indels 35; Gaps 7;

DY 1 TTVEKEPIIL-KDTGEVSELKPHRYVTITONGKEMSTIVSEDFILPYKGLKGYQFD 59
DB 157 TTKTEYAKAKNAVYKAKNAVQKANOVLKAKKASS-----YNYLI----- 197
OY 60 GWEISGF--EGKQDAG-----YVNLSDKTFIKPVFKIEKKKEENKPTDVSKKQNP 112
DB 198 GWEFGGVPPEHKKEEMNLSTLYSSKDKENISKENDVDLDE-KEESELEFTEBELEKNE 256
OY 113 QVNHSQLNESHKEDLQREHSHOKSDSTQVATV-----LDKNNISKSTNN 161
DB 257 ETESEISEDEEBEKEEBEKEENKKEQKQENENNDQKQMEA-----QNLISKQNNN 312

```

```

RESULT 9
O25995 PRELIMINARY; PRT; 354 AA.
AC O25995.
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, last sequence update)
DT 01-OCT-2002 (TReMBLrel. 22, last annotation update)
DE Antigen.
OS Plasmidium falciparum (isolate NF54).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmidium.
OX NCBI_TaxID=5843;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NF54;
RX MEDLINE=95198774; PubMed=7891748;
RA McColl D.J., Silva A., Foley M., Kun J.F., Favaloro J.M.,
  Thompson J.K., Marshall V.M., Coppel R.L., Kemp D.J., Anders R.F.;
RT "Molecular variation in a novel polymorphic antigen associated with
  Plasmidium falciparum merozoites".
RL Møl. Biochem. Parasitol. 68:53-67 (1994).
DR EMBL; L28825; AAC09377.1; -.
SQ SEQUENCE 354 AA; 40119 MW; 3A7256152F48B527 CRC64;

Query Match 13.1%; Score 111.5; DB 5; Length 354;
Best Local Similarity 22.5%; Pred. No. 1.7;
Matches 41; Conservative 34; Mismatches 62; Indels 45; Gaps 7;

DY 1 TTVEKEPIIL-KDTGEVSELKPHRYVTITONGKEMSTIVSEDFILPYKGLKGYQFD 59
DB 154 TTKTEYAKAKNAVYKAKNAVQKANOVLKAKKASS-----YDYLI----- 194
OY 60 GWEISGF--EGKQDAG-----YVNLSDKTFIKPVFKIEKKKE-----EN 99
DB 195 GWEFGGVPPEHKKEEMNLSTLYSSKDKENISKENDVDLDEKEEELTEBELEKNEB 254
OY 100 KPTDVSKKKDPQVNHSQLNESHKEDLQREHSHOKSDSTQVATVATLDKNNISKSTT 159
DB 255 ETESEISEDEEBEKEEBEKEENKKEQKQENENNDQKQMEA-----QNLISKQNN 309
OY 160 NN 161
DB 310 NN 311

RESULT 10
O81J55 PRELIMINARY; PRT; 354 AA.
AC O81J55.
DT 01-MAR-2003 (TReMBLrel. 23, Created)
DT 01-MAR-2003 (TReMBLrel. 23, last sequence update)
DT 01-MAR-2003 (TReMBLrel. 23, last annotation update)
DB Merzoite surface protein 3.
GN PF10_0345.
OS Plasmidium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmidium.
OX NCBI_TaxID=36329;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=3D7;
RX MEDLINE=22255705; PubMed=12368864;
RA Gardner M.J., Hall N., Fung B., White O., Berrihan M., Hyman R.M.,
  Carlson J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,
  Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,
  Chan M.-S., Nene V., Shallow S.J., Suh B., Peterson J., Angiuoli S.,

```



```
QY 11 DTGENSELKPHRYVTVTIONGKMSSTIVSEDFILPYKGELEKGYQ----- 57
DB 167 ETGE-----RNSRNPFYTKTKE-----YAGKVEKDYERANAYOKANQAV 207
QY 58 -----PD-----GWEISGF--BGRDAG-----YYINLSKOFIFPVFKKIEKKEEN 99
DB 208 LKAKKSSYDYLIGWFGGVPFKKKEENMLSHLYSSKOKENISKENDVDLDE-KBEEX 266
QY 100 KPTFDVSKKKNQNVHNSQLN-----ESHKREDLQREHSHOKSDSTKQVAT 146
DB 267 ESTEERELKKEEETRESEISEDEBEKEKEKEENKKGQKEQSENENDOKKME 325
QY 147 VLDKNNISSKSTNN 161
DB 326 ----QWLSKNNNN 336

RESULT 14
QY 0815P3 PRELIMINARY; PRT; 829 AA.
AC 0815P3;
DT 01-MAR-2003 (TRENBLrel. 23, Created)
DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)
DE 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
GN Hypothetical protein.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
ON NCBI_TaxID=36329;
RX MEDLINE=22255705; PubMed=12368864;
RC STRAIN=3D7;
RA Gardner M.J., Hall N., White O., Berriman M., Hyman R.W.,
RA Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,
RA Bisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,
RA Chan M.-S., Nene V., Shallow S.J., Sub B., Peterson J., Angiuoli S.,
RA Perera M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B.,
RA Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,
RA McFadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,
RA Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,
RA Praeger C.M., Barrell B.;
RT "Genome sequence of the human malaria parasite Plasmodium
RT falciparum."
RL Nature 419:498-511(2002).
DR EMBL; AEO14848; AAN36341.1; -.
KW Hypothetical protein.
SQ SEQUENCE 829 AA; 98816 MW; EF26753301B2CE93 CRC64;

Query Match 13.0%; Score 110; DB 5; Length 829;
Best Local Similarity 24.0%; Pred. No. 5.4;
Matches 43; Conservative 39; Mismatches 55; Indels 42; Gaps 9;

QY 2 TYKEFLINDTGEVSLKPHRYVTVTIONGKMSSTIVSEDFILPYKGELEKGYQPDGM 61
DB 491 TSDSLTINKTKNTKI--IQP--LEYLLKNTIGD--KTLMTKRDYLDVPHPMKKYILN-- 542
QY 62 EISGPEKKDAGVYINLSKOTFIKPVFKKIEKKEENKPT-----FDVSK 107
DB 543 -----KKEYLNFESLN-----FRELEKNNKRRKKKGGTHINNKNDAEYMLKTKYK 587
QY 108 KKN--POVHNSQLNESH-KREDLQREHSHOKSDSTKQVATVLDKNNISSKSTNNPNK 164
DB 588 KKNKPNENNTLNDNICKENKLVH--DNSLKQBOIINDKQVIEHTKIYDQK 643

RESULT 15
QY 0812K8 PRELIMINARY; PRT; 609 AA.
AC 0812K8;
DT 01-MAR-2003 (TRENBLrel. 23, Created)
DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)
```

```
DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
DE Large cyclophilin-like protein.
GN PPI1490C.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
ON NCBI_TaxID=36329;
RX MEDLINE=22255708; PubMed=12368867;
RC STRAIN=FC27;
RA Hall N., Pain A., Berriman M., Churcher C., Harris B., Harris D.,
RA Mungall K., Bowman S., Alkin R., Baker S., Barron A., Brooks K.,
RA Bucke C.O., Burrows C., Cherevach I., Chillingworth C.,
RA Chillingworth T., Christodoulou Z., Clark L., Clark R., Corton C.,
RA Cronin A., Davies R., Davis P., Dear P., Dearden F., Doggett J.,
RA Felwell T., Goble A., Goodhead I., Gwilliam R., Hamlin N., Hance Z.,
RA Harper D., Hauser H., Hornsby T., Holroyd S., Horrocks P.,
RA Humphray S., Jørgels K., James K.D., Johnson D., Kethorrou A.,
RA Knights A., Kontorov B., Kyes S., Larke N., Lawson D., Leonard N.,
RA Line A., Maddison M., McLean J., Mooney P., Moulé S., Murphy L.,
RA Oliver K., Ormond D., Price C., Quail M.A., Rabinowitch B.,
RA Rajadream M.A., Rutter S., Rutherford K.M., Sanders M., Simmonds M.,
RA Seeger K., Sharp S., Smith R., Squares R., Squares S., Stevens K.,
RA Taylor K., Tivey A., Unwin L., Whitehead S., Woodward J.,
RA Sulston J.E., Craig A., Newbold C., Barrell B.G.;
RT "Sequence of Plasmodium falciparum chromosomes 1, 3-9 and 13."
RL Nature 419:527-531(2002).
DR EMBL; AL929358; CAD51984.1; -.
SQ SEQUENCE 609 AA; 72552 MW; 8CDF86E85FF9A021 CRC64;

Query Match 12.9%; Score 109.5; DB 5; Length 609;
Best Local Similarity 23.9%; Pred. No. 4.2;
Matches 42; Conservative 36; Mismatches 79; Indels 19; Gaps 5;

QY 5 EFLINDTGEVSLKPHRYVTVTIONGKMSSTIVSEDFIL--PYKGELEKGYQ 57
DB 114 KILINKTKTGVNYSSEIKL--FHNKNEFDPDTHKISKEDPIVLDPLNKTSTVASH 170
QY 58 FEGWEISGREGKADAGVYINLSKOTFIKPVFKKIEKKEENKPTDVSKKKNP----- 112
DB 171 LMDRDQEDNEKKKKKNIFAPSIQDNGAIMSILKEMERAKKKEKKEQEEBEKKKKYGI 230
QY 113 QVNSQLNSHAKREDLQREHSHOKSDSTKQVATVLD---KNISSKSTNNPNK 164
DB 231 QVNSPSNSESSEHSEKKEKWLKNKQKQYNNKIHSTYDSSENEIKIKCTNYSN 286

RESULT 16
QY 026019 PRELIMINARY; PRT; 380 AA.
AC 026019;
DT 01-NOV-1996 (TRENBLrel. 01, Created)
DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)
DT 01-OCT-2002 (TRENBLrel. 22, Last annotation update)
DB Polymorphic antigen precursor.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
ON NCBI_TaxID=5833;
RX MEDLINE=95198774; PubMed=7891748;
RC STRAIN=FC27;
RA McColl D.J., Silva A., Foley M., Kun J.Z., Favaloro J.M.,
RA Thompson J.K., Marshall V.M., Coppel R.L., Kemp D.J., Anders R.F.;
RT "Molecular variation in a novel polymorphic antigen associated with
RT Plasmodium falciparum merozoites."
RL Mol. Biochem. Parasitol. 68:53-67(1994).
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FC27;
RX MEDLINE=98156743; PubMed=9497029;
RA McColl D.J., Anders R.F.;
RT "Conservation of structural motifs and antigenic diversity in the
RT Plasmodium falciparum merozoite surface protein-3 (MSP-3)."
```

RL Mol. Biochem. Parasitol. 90:21-31(1997).  
 DR EMBL, L07944; AAC09378.1; -.  
 KW Signal.  
 FT SIGNAL 1 25 POTENTIAL.  
 FT CHAIN 26 380 POLYMORPHIC ANTIGEN.  
 SQ SEQUENCE 380 AA; 43290 MW; 0986CA13930944C2 CRC64;

Query Match 12.9%; Score 109; DB 5; Length 380;

Best local similarity 23.5%; Pred. No. 2.7;  
Matches 42; Conservative 34; Mismatches 67; Indels 36; Gaps 8;

QY 12 TGEVSHLKPRVTVTTIQNGKMSSTTVSEDFILPYV-----KGELEKGYQFD-GWE 62  
 DB 166 TSETPR-KGSRINLPSKRTKEVAKKANAYEKANAYOKANQAVLKAKEASDYILGWE 224  
 QY 63 ISGF--EGKKDAG-----YINLSKOTFTFVFKKIHKEKEENKPTFDVSKKKNPQN 115  
 DB 225 FGCGVPEHKKEENMLSHLYSSSKDKENISKENDVDLDE-KBEAEETEEBEKEKEEET 283  
 QY 116 HSGQLN-----ESHKEDLQREHHSOKSDSTKVTVTLVDKNNISKSTNN 161  
 DB 284 ESEISDEEBEKEEKEEKEEKEEKEEKEEKEEKEEKEEKEEKEEKEEKEEKEEKEEKEE 337

## RESULT 17

Q81436 PRELIMINARY; PRT; 3008 AA.  
 ID 081436;  
 AC 081436;  
 DT 01-MAR-2003 (Tremblrel. 23, Created)  
 DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)  
 DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)  
 DE Hypothetical protein.  
 GN PFE0325W  
 OS Plasmodium falciparum (isolate 3D7).  
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
 OC NCBI\_TaxID=36329;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Devlin K., Baker S., Davies P., Mungal K., Berriman M., Pain A.,  
 RA Hall N., Bowman S., Churcher C., Quail M., Barrett B.;  
 RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.

RX MEDLINE=2255706; PubMed=1236867;  
 RA Hall N., Pain A., Berriman M., Churcher C., Harris B., Harris D.,  
 RA Mungall K., Bowman S., Atkin R., Baker S., Barron A., Brooke K.,  
 RA Bucke C.O., Burrows C., Cherevach I., Chillingworth C.,  
 RA Chillingworth T., Christodoulou Z., Clark L., Clark R., Corson C.,  
 RA Cronin A., Davies R., Davis P., Dear P., Dearden F., Doggett J.,  
 RA Felwell T., Goble A., Goodhead I., Gwilliam R., Hamlin N., Hance Z.,  
 RA Harper D., Hauser H., Hornsby T., Holroyd S., Horrocks P.,  
 RA Humphray S., Jagele K., James K.D., Johnson D., Kerhornou A.,  
 RA Knights A., Konfortov B., Kyes S., Larke N., Lawson D., Lennard N.,  
 RA Line A., Maddison M., McLean J., Moorey P., Moule S., Murphy L.,  
 RA Oliver K., Ormond D., Price C., Quail M.A., Rabinowitsch E.,  
 RA Rajandream M.A., Rutter S., Rutherford K.M., Sanders M., Simmonds M.,  
 RA Seeger K., Sharp S., Smith R., Squares R., Squares S., Stevens K.,  
 RA Taylor K., Tivey A., Unwin L., Whitehead S., Woodward J.,  
 RA Sulston J.E., Craig A., Newbold C., Barrett B.G;  
 RT "Sequence of Plasmodium falciparum chromosomes 1, 3-9 and 13.";  
 RL Nature 419:527-531(2002).  
 DR EMBL, AL929351; CAD51431.1; -.  
 KW Hypothetical protein.  
 SQ SEQUENCE 3008 AA; 356023 MW; 60BCBBEELSC59984 CRC64;

Query Match 12.9%; Score 109; DB 5; Length 3008;

Best local similarity 32.4%; Pred. No. 26;  
Matches 35; Conservative 22; Mismatches 31; Indels 20; Gaps 6;

QY 74 YVINLSK-----DTPIKVPFKTIKKEEENKPTFDVSKKKNPQNPNVNSQ---LNESSHKE 126  
 DB 2310 YDIELSKIEKGAISGIPFTD-EKKKEEKKK--EVNKKKEEKKKEEKKKEEKKKEEKKKE 2366

QY 127 DLQREH-----SQKSDSTKVTVTLVDKNNISK-----STNNPNK 164  
 DB 2367 ENKKEEKKKEEKKKEEKKKEEKKKEEKKKEEKKKEEKKKEEKKKEEKKKEEKKKEE 2414

## RESULT 18

ID 077355 PRELIMINARY; PRT; 600 AA.  
 AC 077355;  
 DT 01-NOV-1998 (Tremblrel. 08, Created)  
 DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)  
 DT 01-OCT-2002 (Tremblrel. 22, Last annotation update)  
 DE Hypothetical 71.7 kDa protein.  
 GN PFC0465C, MAL3P4.20.  
 OS Plasmodium falciparum (isolate 3D7).  
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
 OC NCBI\_TaxID=36329;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Bowman S., Lawson D., Baeham D., Brown D., Chillingworth T.,  
 RA Churcher C.M., Craig A., Davies R.M., Devlin K., Felwell T.,  
 RA Gentles S., Gwilliam R., Hamlin N., Harris D., Holroyd S., Hornsby T.,  
 RA Horrocks P., Jagele K., Jassal B., Kyes S., McLean J., Moule S.,  
 RA Mungall K., Murphy L., Oliver K., Quail M.A., Rajandream M.-A.,  
 RA Rutter S., Skelton J., Squares R., Squares S., Sulston J.E.,  
 RA Whitehead S., Woodward J.R., Newbold C., Barrett B.G.;  
 RT "The complete nucleotide sequence of chromosome 3 of Plasmodium falciparum.";  
 RL Nature 400:532-538(1999).  
 DR EMBL, AL008970; CA015610.2; -.  
 DR InterPro; IPR002483; PFI.  
 DR Pfam; PF01480; PFI; 1.  
 DR SMART; SM00311; PFI; 1.  
 KW Hypothetical protein.  
 SQ SEQUENCE 600 AA; 71663 MW; 57BAB42565CAD64C CRC64;

Query Match 12.8%; Score 108.5; DB 5; Length 600;  
 Best local similarity 29.3%; Pred. No. 4.9;  
 Matches 54; Conservative 22; Mismatches 47; Indels 61; Gaps 12;

QY 37 IVSEEDFILPYV-----KGELEKGYQFDGWEISGFEKCK---DAGVYINLSKOTFTKPV 87  
 DB 60 ILGFEDILIVCYCISQKQSEKCK---DGEEDKYINAKKLINLTGFIQNKSDIFIEL 116  
 QY 88 FKKI-EKKKEE-----ENKPTFDVSK-KKNPQVNSQLNF-----SHRK 125  
 DB 117 LELLINEKKEEHIADTLANEK-TNDIKVKNENENINENYNNKDISKDKERHVSQN 175  
 QY 126 E-----DLQREH-----SQKSDSTKVTVTLVDKNNISKSTNN 160  
 DB 176 EHNINNVNLKKEKETTDIQRKRGKSLSQKSDSYKRRPNPKRTSTIER-SLSNRYDE 234  
 QY 161 NPNK 164  
 DB 235 KTNK 238

## RESULT 19

Q9FUK9 PRELIMINARY; PRT; 470 AA.  
 ID Q9FUK9;  
 AC Q9FUK9;  
 DT 01-MAR-2001 (Tremblrel. 16, Created)  
 DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)  
 DT 01-OCT-2002 (Tremblrel. 22, Last annotation update)  
 DE Gb|AF20218.1.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC Eustrodi II; Brassicales; Brassicaceae; Arabidopsis.  
 OC NCBI\_TaxID=3702;  
 RN [1]

RP SEQUENCE FROM N.A.  
RC STRAIN=Columbia;  
RX MEDLINE=9087489; PubMed=9872454;  
RA Nakamura Y., Sato S., Asamizu E., Kaneko T., Kotani H., Miyajima N.,  
RT Tabata S.;  
RT "Structural analysis of Arabidopsis thaliana chromosome 5. VII.  
RT Sequence features of the regions of 1,013,767 bp covered by sixteen  
RT physically assigned P1 and TAC clones."  
RL DNA Reg. 5:297-308(1998).  
DR EMBL: AB015468; BAB10694.1; -  
SQ SEQUENCE 470 AA; 53758 MW; 6D686CE72B35AC54 CRC64;

Query Match 12.7%; Score 107.5; DB 10; Length 470;  
Best Local Similarity 20.1%; Pred. No. 4.4;  
Matches 36; Conservative 32; Mismatches 66; Indels 45; Gaps 5;

QY 19 KHRVTVTIONGKMSSTVSEEDFLLPVYKGLKELGYOPDGEISGFE-----GKK 70  
DB 82 RNRVTVDVQNNNSNGSK-----YVDLARRIKIDE-EATGSGAQRIDHPNOK 129  
QY 71 DAGVYINLSKOTFFIKVFKKIEKKEENKPTFDVSKKDN----- 111  
DB 130 NVGITEKAFENSPIESTSHRVDNKKINQKXFTAKSSENAVSRVSPGADHRAVGMK 189  
QY 112 PQVNSQLNE-----SHRKEDLQREHSGOKSDSTQVATVLDKNNISSKSTNNPNK 164  
DB 190 PHEMRDQVQTSASEKSHKENVTKSEKRPDQGVKKTAKDKDRNKEKKEETESINK 248

## RESULT 20

096229 PRELIMINARY; PRT; 951 AA.

ID 096229  
AC 096229; MEDLINE=99021743; PubMed=9804551;  
DT 01-MAY-1999 (TrEMBLrel. 10, Created)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE Hypothetical protein.  
GN PF80680W.  
OS Plasmodium falciparum (isolate 3D7).  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.  
OX NCBI\_TaxID=36329;  
RN (1)  
RP SEQUENCE FROM N.A.  
RC STRAIN=3D7;  
RX MEDLINE=99021743; PubMed=9804551;  
RA Gardner M.J., Tetcelin H., Carucci D.J., Cummings L.M., Aravind L.,  
RA Koonin E.V., Shallom S., Mason T., Yu K., Fujii C., Petersen J.,  
RA Shen K., Jing J., Astor C., Lai Z., Schwartz D.C., Petrea M.,  
RA Salberg S., Zhou L., Sutton G.G., Clayton R., White O., Smith H.O.,  
RA Praeger C.M., Adams M.D., Venter J.C., Hoffman S.L.;  
RT "Chromosome 2 sequence of the human malaria parasite Plasmodium  
RT falciparum."  
RL Science 282:1126-1132(1998).  
RN (2)  
RP SEQUENCE FROM N.A.

RC STRAIN=3D7;  
RX MEDLINE=2255705; PubMed=1236864;  
RA Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W.,  
RA Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,  
RA Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,  
RA Chan M.-S., Nene V., Shallom S.J., Suh B., Peterson J., Angiuoli S.,  
RA Petrea M., Allen J., Selengut J., Haft D., Mather M.W., Valdivia A.B.,  
RA Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,  
RA McPadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,  
RA Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,  
RA Praeger C.M., Bartell B.;  
RT "Genome sequence of the human malaria parasite Plasmodium  
RT falciparum."  
RL Nature 419:498-511(2002).  
DR EMBL: AB001410; AAC71925.2; -  
KW Hypothetical protein.  
SQ SEQUENCE 951 AA; 112486 MW; AC8D889358A84F4F CRC64;

Query Match 12.5%; Score 106; DB 5; Length 951;  
Best Local Similarity 24.3%; Pred. No. 12;  
Matches 42; Conservative 35; Mismatches 40; Indels 56; Gaps 10;

QY 10 KOTGEVSELKPRVT-VTTIONGKMSSTVSEEDFLLPVYKGLKELGYOPDGEI--SGF 66  
DB 128 EKKNKINKDLHRONELNLSGK-----NQDI-----NNEKGGKQ--DISNSNA 170  
QY 67 EKKKQAGVYINLSKOTFFIKVFKKIEKKE-----EENKPTFD-----VSKKQNP 112  
DB 171 ENKXD-----VREGVYELKKEKKEKESIDDKHVEBKSSDKHVEBKSSDDH 218  
QY 113 QVNSQLNE-----SHRKEDLQREHSGOKSDSTQVATVLDKNNISSKSTNNPNK 164  
DB 219 KVENKKSDDHKIEVYKVEHBEDEE-----DIKESKSNKNQDNK 262

## RESULT 21

09NFV9 PRELIMINARY; PRT; 329 AA.

ID 09NFV9  
AC 09NFV9; MEDLINE=20416497; PubMed=10960178;  
DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)  
DE Merozoite surface protein 3 (Fragment).  
GN MSP3.  
OS Plasmodium falciparum (isolate 7G8).  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.  
OX NCBI\_TaxID=57266;  
RN (1)  
RP SEQUENCE FROM N.A.  
RC STRAIN=7G8;  
RX MEDLINE=20416497; PubMed=10960178;  
RA Ocken D.M.N., Thomas A.W., Conway D.J.;  
RT "Allelic lineages of the merozoite surface protein 3 gene in  
RT Plasmodium reichenowi and Plasmodium falciparum."  
RL Mol. Biochem. Parasitol. 109:185-188(2000).  
DR EMBL: AJ252287; CAB85901.1; -  
FT NON\_TER 1  
FT NON\_TER 329  
FT NON\_TER 329  
SQ SEQUENCE 329 AA; 36916 MW; C5B045DB5E21A159 CRC64;

Query Match 12.3%; Score 104.5; DB 5; Length 329;  
Best Local Similarity 24.3%; Pred. No. 4.9;  
Matches 43; Conservative 30; Mismatches 63; Indels 41; Gaps 8;

QY 1 TTVEKPIFN-KDTGEVSELKPRVT-VTTIONGKMSSTVSEEDFLLPVYKGLKELGYOPD 59  
DB 144 TKTGEYAEKQKAYBEAKQAYOKANQAVLKAKKASS-----YDIL----- 164  
QY 60 GWETSGF--EGKKDAG-----YVINLSKOTFFIKVFKKIEKKEENKPTFDVSKKQNP 112  
DB 185 GWETSGGVFEHKKEKEMLSHLYSSKDKENISKENDVIDE-KEEBAEETEEBELBEKNE 243  
QY 113 QVNSQLNE-----SHRKEDLQREHSGOKSDSTQVATVLDKNNISSKSTNNPNK 161  
DB 244 BETESKISDBEEBEKEBEKEBEKEQAKESNENDQCKDMA-----QNLISKQNNN 295

## RESULT 22

094C59 PRELIMINARY; PRT; 540 AA.

ID 094C59  
AC 094C59; MEDLINE=2001 (TrEMBLrel. 19, Created)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE Putative phosphatidyl-inositol-transfer protein.  
GN T518.14.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
OX NCBI\_TaxID=3702;

```

RN [1]
RP SEQUENCE FROM N.A.
RA Yamada K., Liu S.X., Sakano H., Pham P.K., Banh J., Chung M.K.,
RA Goldsmith A.D., Lee J.M., Quach H.L., Tang C., Toriumi M., Yu G.,
RA Bowser L., Carninci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J.,
RA Jones T., Kamiya A., Karlin-Neumann G., Kawai J., Kim C., Koeseima E.,
RA Lam B., Lin J., Meyers M.C., Miranda M., Narusaka M., Nguyen M.,
RA Palm C.J., Sakurai T., Satou M., Seki M., Shinn P., Southwick A.,
RA Shinohara K., Davis R.W., Becker J.R., Theologis A.,
RT "Full Length cDNA of gene T518.14 (GI:4587525).";
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AY05162; AAK59666.1; -
DR InterPro: IPR001251; CRAL_TRIO.
DR InterPro: IPR001071; RetBind/locTrans.
DR Pfam: PF00650; CRAL_TRIO.1.
DR Pfam: PF03765; CRAL_TRIO_N.1.
DR PRINTS: PR00180; CRETINALDHP.
DR SMART: SM00516; SEC14.1.
DR PROSITE: PS50191; CRAL_TRIO.1.
SQ SEQUENCE 540 AA; 6116 MW; 0C2592EB18ACFAC6 CRC64;

Query Match 12.3%; Score 104; DB 10; Length 540;
Best Local Similarity 25.3%; Pred. No. 9.2;
Matches 40; Conservative 28; Mismatches 64; Indels 26; Gaps 6;

QY 1 TTVEFLINKDTGVSELPKPRVVTYIIONGKMSSTIYSEDFILPYKGELEKGYQFDG 60
DB 27 TTVA-VVEETKVEDESKP-----EGVEKSNASFEESDFPADLSESK----- 70
QY 61 WEISGFGKDGAGVYNLSKDTFIKPVFK--IEEKEENKPTDVSKKKNQVNH 117
DB 71 -ALSDLSKLEBAVDN---TLTKTKKESPMKKEEVVKEAEVKKKE--EAAEE 123
QY 118 QLNESHKREDLQREHSQKSDSTVDYATVLDKNISS 155
DB 124 KVEBKSEAVVTEAPKAETVBAVTEIIPKEVTT 161

RESULT 23
Q9SA84 PRELIMINARY; PRT; 540 AA.
ID Q9SA84;
AC Q9SA84;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
DE T518.14 protein (Hypothetical protein).
GN T518.14 OR ATIG30690.
OS Arabidopsis thaliana (mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RA Vysotskaya V.S., Schwartz J.R., Yu G., Toriumi M., Lenz C., Liu S.,
RA Li J., Kremenetskaia I., Luros J., Ngan I., Gonzalez A., Alcafi H.,
RA Araujo R., Chao Q., Conn L., Conway A.B., Dunn P., Hansen N.,
RA Huzar L., Kim C., Palm C., Rowley D., Shinn P., Walker M.,
RA Davis R.W., Becker J.R., Federicel N.A., Theologis A.,
RT "Arabidopsis thaliana chromosome 1 BAC T518 sequence.";
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Yamada K., Chan M., Chang C.H., Dale J.M., Hsuan V.W., Lee J.M.,
RA Quach H.L., Tang C.C., Toriumi M., Wallender E.K., Wong C., Wu H.C.,
RA Yu G., Yuan S., Chen H., Cheuk R., Jones T., Kim C.J., Nguyen M.,
RA Palm C.J., Shinn P., Southwick A., Tripp M.G., Wu T., Davis R.W.,
RA Becker J.R., Theologis A.,
RT "Arabidopsis Open Reading Frame (ORF) Clones.";
RL Submitted (OCT-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL: AC007060; AAD25756.1; -
DR EMBL: BT000959; AAN41359.1; -

```

```

DR InterPro: IPR001251; CRAL_TRIO.
DR InterPro: IPR001071; RetBind/locTrans.
DR Pfam: PF00650; CRAL_TRIO.1.
DR Pfam: PF03765; CRAL_TRIO_N.1.
DR PRINTS: PR00180; CRETINALDHP.
DR SMART: SM00516; SEC14.1.
DR PROSITE: PS50191; CRAL_TRIO.1.
KV Hypothetical protein.
SQ SEQUENCE 540 AA; 61189 MW; 0C2590D518ACFB58 CRC64;

Query Match 12.3%; Score 104; DB 10; Length 540;
Best Local Similarity 25.3%; Pred. No. 9.2;
Matches 40; Conservative 28; Mismatches 64; Indels 26; Gaps 6;

QY 1 TTVEFLINKDTGVSELPKPRVVTYIIONGKMSSTIYSEDFILPYKGELEKGYQFDG 60
DB 27 TTVA-VVEETKVEDESKP-----EGVEKSNASFEESDFPADLSESK----- 70
QY 61 WEISGFGKDGAGVYNLSKDTFIKPVFK--IEEKEENKPTDVSKKKNQVNH 117
DB 71 -ALSDLSKLEBAVDN---TLTKTKKESPMKKEEVVKEAEVKKKE--EAAEE 123
QY 118 QLNESHKREDLQREHSQKSDSTVDYATVLDKNISS 155
DB 124 KVEBKSEAVVTEAPKAETVBAVTEIIPKEVTT 161

RESULT 24
Q44016 PRELIMINARY; PRT; 325 AA.
ID Q44016;
AC Q44016;
DT 01-JUN-1998 (Tremblrel. 06, Created)
DT 01-JUN-1998 (Tremblrel. 06, Last sequence update)
DT 01-OCT-2002 (Tremblrel. 22, Last annotation update)
DE G5 ORF.
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.
OX NCBI_TaxID=44689;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NC4;
RX MEDLINE=96198836; PubMed=9539429;
RA Riben W.K. Jr., Gonzales C.M., Gonzales S.T., Pilkington K.J.,
RA Kiyosawa H., Hughes J.E., Welker D.L.,
RT "Dictyostelium discoideum nuclear plasmid Ddps is a chimera related to
the Ddpl and Ddp2 plasmid families.";
RL Genetics 148:1117-1125(1998).
DR EMBL: U00796; AAC18634.1; -
SQ SEQUENCE 325 AA; 38448 MW; 69443D0C632058A6 CRC64;

Query Match 12.2%; Score 103.5; DB 5; Length 325;
Best Local Similarity 23.8%; Pred. No. 5.7;
Matches 43; Conservative 25; Mismatches 50; Indels 63; Gaps 9;

QY 21 HRVVTIONGKMSSTIYSEDFILPYK-GEL--EKGYQDGMHISGFGK----- 69
DB 57 HRTTISIGN-RFSYKXIGDEEKLPRISKGBLYLNLLEDFNFIK--EGHLRKSQMF 112
QY 70 ---KDAGV-----INLSKDTFIKPV-----FKK----- 90
DB 113 NHIKDSGYVATNEIBIFLESCTLCKEITAOIKRNSYKRNINLPEREEREEEREE 172
QY 91 --IEEKEENKPTDVSKKKNQVNHQLNESIRK-----DLQREHSQKSDSTK 141
DB 173 EEEBQEEVEVKPTISBEEBETPAVSEBKEEBEETPAVSEBKEEBEEDK 232
QY 142 D 142
DB 233 D 233

RESULT 25
Q9LH98

```



```

ID 09LH98 PRELIMINARY; PRT; 2081 AA.
AC 09LH98;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, last annotation update)
DE Genomic DNA, chromosome 3, BAC clone: T19N8.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC euroside II; Brassicales; Brassicaceae; Arabidopsi.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Columbia;
RA Kaneko T., Kato T., Sato S., Nakamura Y., Asamizu R., Tabata S.;
RL Submitted (May-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Columbia;
RA MEDLINE=2036309; PubMed=10907853;
RX Nakamura Y.;
RA "Structural analysis of Arabidopsis thaliana chromosome 3. II.
RT Sequence features of the regions of 4,251,695 bp covered by ninety pl,
RT TAC and BAC clones."
RL DNA Res. 7:217-221 (2000).
DR EMBL; AF002057; BAB03174.1; -
SQ SEQUENCE 2081 AA; 232851 MW; D3603E1F85EFF29 CRC64;

Query Match
Best Local Similarity 12.1%; Score 103; DB 10; Length 2081;
Matches 46; Conservative 29; Mismatches 72; Indels 30; Gaps 7;

OY 7 ILNKDGVSEELK---PHRTVTITONGKEMSTIVSEDFILPVYKGLKGYOPDGMW 63
DB 1639 IYERNGKEDSLKESSEDKTYIRINGBELSTEEGKD-----GKIEEGK--EGKN 1688
OY 64 SGFEKKD-----AGVINISKDTFIKPVFKIEKKKEENK--TFDVSK-KKD 110
DB 1689 STKESSKDKIEGMEGKENSSTKESKDKINEHDKATMEGSKDGJNSTGDSKD 1748
OY 111 NPQVNSQLNESHKREDLQ---REHSQKSDSTKQVATATVLDKNNISKSTNNPN 163
DB 1749 SKSVINGVKDLSLKDSSKNGDINETNGKEDSVKDNVTIEIQNDNLSLTNSSEPN 1805

RESULT 26
O81226 PRELIMINARY; PRT; 238 AA.
AC 081226;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, last annotation update)
DE Hypothetical protein.
GN PF10765W.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=22255708; PubMed=12368867;
RA Hall N., Pain A., Berriman M., Churcher C., Harris B., Harris D.,
RA Mungall K., Bowman S., Atkin R., Baker S., Barron A., Brooks K.,
RA Buceae C.O., Burrows C., Chervach I., Chillingworth C.,
RA Chillingworth T., Christodoulou Z., Clark L., Clark R., Corton C.,
RA Crotin A., Davies R., Davis P., Dear P., Dearden F., Doggett J.,
RA Feltwell T., Goble A., Goodhead I., Gwilliam R., Hamlin N., Hance Z.,
RA Harper D., Hauser H., Hornsby T., Holroyd S., Horrocks P.,
RA Humphrey S., Jagels K., James K.D., Johnson D., Kerhornou A.,
RA Knights A., Konfortov B., Kyes S., Larke N., Lawson D., Leonard N.,
RA Line A., Maddison M., McLean J., Mooney P., Moulie S., Murphy L.,
RA Oliver K., Ormond D., Price C., Quail M.A., Rabinowitsch E.,
RA Rajandream M.A., Rutter S., Rutherford K.M., Sanders M., Simmonds M.,
RA Seeger K., Sharp S., Smith R., Squares R., Squares S., Stevens K.,

```

```

RA Taylor K., Tivey A., Unwin L., Whitehead S., Woodward J.,
RA Sulzeron J.B., Craig A., Newbold C., Barrell B.G;
RT "Sequence of Plasmodium falciparum chromosomes 1, 3-9 and 13."
RL Nature 419:527-531 (2002).
DR EMBL; AL929356; CAD51839.1; -
KW Hypothetical protein.
SQ SEQUENCE 238 AA; 28481 MW; DA175A2B5A109B49 CRC64;

Query Match
Best Local Similarity 12.1%; Score 102.5; DB 5; Length 238;
Matches 49; Conservative 26; Mismatches 65; Indels 35; Gaps 10;

OY 10 KOTGEVSELKPHRTVTITONGKEMSTIVSEDFILPVYKGLKGYOPDGMWISG---F 66
DB 38 KEKERTQETFKKD-----ENNKSNNNVNVNSENF--WKIYKNILKQYHVEKFKQENITE 91
OY 67 EKKKA-----GYVINISKDTFIKPVFKIEKKKEENKPTFDVSKKDN--PQV--N 115
DB 92 ESKKATSRKKIKQNFKLKKNLLRKVLK--KAPLAKATVPSLKKTYNVPPIH SNN 148
OY 116 H--SQLNESHKRED-----LOREHSQKSDSTKQVATATVLDKNNISKSTNNPN 163
DB 149 HNISQDNHKKQDKNKKIKLLIKLSKNDNN-----NINNKNNQKCTTKLN 198

RESULT 27
O9PPL5 PRELIMINARY; PRT; 312 AA.
AC 09PPL5;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, last annotation update)
DE Putative membrane protein.
GN C30692C.
OS Campylobacter jejuni.
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;
OC Campylobacteraceae; Campylobacter.
OX NCBI_TaxID=197;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NCTC 11168;
RX MEDLINE=20150912; PubMed=10688204;
RA Parish J., Wren B.W., Mungall K., Ketley J.M., Churcher C.,
RA Basham D., Chillingworth T., Davies R.M., Feltwell T., Holroyd S.,
RA Jagels K., Karlyshev A.V., Moulie S., Pallen M.J., Penn C.W.,
RA Quail M.A., Rajandream M.A., Rutherford K.M., van Vliet A.H.M.,
RA Whitehead S., Barrell B.G.;
RT "The genome sequence of the food-borne pathogen Campylobacter jejuni
RT reveals hypervariable sequences."
RL Nature 403:665-668 (2000).
DR EMBL; AL139076; CAB72966.1; -
KW Complete proteome.
SQ SEQUENCE 312 AA; 37221 MW; 0004FA7836A741B8 CRC64;

Query Match
Best Local Similarity 12.0%; Score 101.5; DB 16; Length 312;
Matches 46; Conservative 34; Mismatches 73; Indels 31; Gaps 9;

OY 5 EFLNKDGVSEELKPHRTVTITONGKEMSTIVSEDFILPVYK-----GELE 53
DB 5 DIKQRQNIQRKQFLFSR--AINQKPLDDLRBEISSDDLIRRRFKKTPKPLEED 61
OY 54 KGYOPDGMWISGFEKGKQAGVYNL---SKDTFIKPVFKIEKKKEENKPT--PDVSK 107
DB 62 EYKSKHTKKSNIYTKED--LINVYLEKQSLAKKIFSKMKERREKREKTKKQNF 118
OY 108 KKQNP-----QVNSQLNESHKREDLQREHSQKSDSTKQV--TATVLDKNNISK--STT 159
DB 119 KKAHEKNIQTQTQTSNQAATQTQKKEKELTNSIEKIQTEFTIQPLIEKKLDVK 178
OY 160 NNPN 163
DB 179 NQPN 182

```

## RESULT 28

ID Q9V7J0 PRELIMINARY; PRT; 382 AA.  
 AC Q9V7J0: Q9G081;  
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)  
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)  
 DE C8421 protein (Aspartyl beta-hydroxylase variant 2).  
 GN ASPH OR CG8421 OR CG18658.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Insecta; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Ephemeroptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC NCBI\_TaxID=7227;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celniker S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA Ananides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Mortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champagne M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abiri J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Bernick J., Borkstein P., Borkstein P.,  
 RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke Z., Davenport L.B., Davies P.,  
 RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunlov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferrara C., Ferreira S., Fleischmann W.,  
 RA Flesher C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasner K.,  
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibbagum C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Kechum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Laeko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei V., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacleb J.M.,  
 RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Rainer C., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Klamor I., Simpson M., Skupski M.P., Smith T.,  
 RA Spletter E., Spradling A.C., Stapleton M., Strong R., Sun B.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissenbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of Drosophila melanogaster."  
 RL Science 287:2185-2195(2000).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Celniker S.E., Adams M.D., Krommiller B., Wan K.H., Holt R.A.,  
 RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,  
 RA Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,  
 RA Carlson J.W., Center A., Champagne M., Davenport L.B., Dietz S.M.,  
 RA Dodson K., Dorett V., Doup L.B., Doyle C., Dresnak D., Fartan D.,  
 RA Ferrara S., Frise E., Galle R.F., Garg N.S., George R.A.,  
 RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,  
 RA Ibbagum C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,  
 RA McIntosh T.C., Moy M., Murphy B., Murphy L., Nelson K.A., Nuno J.,  
 RA Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B.,  
 RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,  
 RA Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,  
 RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;

RT "Sequencing of Drosophila melanogaster genome."  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RA Misra S., Crosby M.A., Mathews B.B., Bayraktaroglu L., Campbell K.,  
 RA Hradecky P., Huang Y., Kaminker J.S., Prochnik S.B., Smith C.D.,  
 RA Tupy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.B.,  
 RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,  
 RA Krommiller B., Marshall B., Millburn G., Richter J., Russo S.,  
 RA Searle S.M.J., Smith E., Shu S., Smurnick F., Whitfield B.,  
 RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;  
 RT "Annotation of Drosophila melanogaster genome."  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RA Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter J.C.;  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RA FlyBase;  
 RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.  
 RN [6]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20564328; PubMed=10956655;  
 RA Dinchuk U.E., Henderson N.L., Burn T.C., Huber R., Ho S.P., Link J.,  
 RA O'Neill K.T., Focht R.J., Scully M.S., Hollis J.M., Hollis G.F.,  
 RA Friedmann P.A.;  
 RT "Aspartyl beta-Hydroxylase (Asph) and an evolutionarily conserved  
 RT isofom of Asph missing the catalytic domain share exons with  
 RT Junction."  
 RL J. Biol. Chem. 275:39543-39554(2000).  
 DR EMBL: AE003808; AAF58063.2; -  
 DR EMBL: AF289494; AAC40807.1; -  
 DR FlyBase: FBgn0034075; Asph.  
 SQ SEQUENCE 382 AA; 43287 MW; 60B5C03ABFCEB8 CRC64;  
 Query Match 12.0%; Score 101.5; DB 5; Length 382;  
 Best Local Similarity 24.5%; Pred. No. 9.5;  
 Matches 34; Conservative 29; Mismatches 51; Indels 25; Gaps 5;  
 QY 41 EDPLPVYKGLKGYOPDGM-----SIGSGCKKQAGYI-----NLKDTPIK 85  
 DB 78 EDLDTPLSESRFSK--VFQGVNDEHRDHDGHDVPESGALDDHDDHDDHDEDEDE 135  
 QY 86 PVFKRIEKKKEEKPT-----PDVSKKQNPQVNSQLNSHAKEDLQREHSQKSDS 139  
 DB 136 PLTEBELBELLEEFEPTEDEPPADEYEDDEENNA--GENTATADAEEREBEEDND 193  
 QY 140 TKDYATVYLDKNNTSST 158  
 DB 194 EGTVEATVEATTEATTEAT 212  
 RESULT 29  
 Q9V7I9 PRELIMINARY; PRT; 556 AA.  
 AC Q9V7I9;  
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE CG8421 protein.  
 GN ASPH OR CG8421 OR CG18658.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Emydroidea; Drosophilidae; Drosophila.  
 OC NCBI\_TaxID=7227;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,

SQ	SEQUENCE	556 AA;	63144 MM;	B420980CBDC6357A CRC64:
DR	FlyBase; FBgn0034075; Asph.			
Dt	Query Match	12.0%;	Score 101.5;	DB 5; Length 556;
Df	Best Local Similarity	24.5%;	Pred. No. 14;	
Matches	Matches 34; Conservative 29; Mismatches 51; Indels 25; Gaps 5;			

Oy	41 EDLFLPYKXELGNGYFDGM-----EIGFEEKDAGVI-----NLSKPFIK 85
Dp	78 EDDLTPLESRRFSK-VFFDMVDEHRDHGHGVSPSGALDDHDERHDHHEDER 135
Oy	86 PVFKIKIEKKKEENKPT-----FDVSKKDNDPQNVSQSINSHRKXDLOREHSOKSDS 139
Dp	136 PLTBELEEBEEBEEPFEEDPEADAEYEDEDENNNA-GENITADABEBEBEENDD 193
Oy	140 TKDVATVLDKNNISRSKT 158
Dp	194 EGVEATVEATTEATTREAT 212

RESULT 30  
O9G082 PRELIMINARY; PRT; 785 AA.  
ID O9G082;  
AC O9G082;  
DT 01-MAR-2001 (TREMBREL, 16, Created)  
DT 01-MAR-2001 (TREMBREL, 16, Last sequence update)  
DT 01-MAR-2003 (TREMBREL, 23, Last annotation update)  
DE Asparyl beta-Hydroxylase variant 1 (CG8421-P).  
GN ASPH OR CG8421 OR CG18658.  
OS Drosophila melanogaster (fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
CC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
CX Empidoidea; Empididae; Drosophilinae; Drosophila.  
XX NCBI TaxId=7227;  
[1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=20564328; PubMed=10956665;  
RA Dirichk J.E., Henderson N.L., Burn T.C., Huber R., Ho S.P., Link J.,  
RA O'Neil K.T., Poeth R.D., Scully M.S., Hollis J.M., Hollis G.F.,  
RA Friedman P.A.;  
RT "Aspartyl beta -Hydroxylase (ASPH) domain and an Evolutionarily Conserved  
RT Isoform of ASph Missing the Catalytic Domain Share Exons with  
RT Uncutin".  
RL RT Biol. Chem. 275:39543-39554(2000).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Berkely;  
RM MEDLINE=20196006; PubMed=10731132;  
RX Adams M.J., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Amanatides F.G., Scherer S.E., Li P.W., Hopkins R.A., Galie R.F.,  
RA George R.A., Lewis S.B., Richards S., Ashburner M., Henderson S.N.,  
RA Stutten G.G., Wortman W.R., Yandell M.D., Zhang Q., Chen L.X.,  
BA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,  
BA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
BA Abil U.F., Aghayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
BA Ballen R.M., Basu A., Bakendale J., Bayraktaroglu I., Beasley E.M.,  
BA Beeson K.Y., Benos P.V., Bernan B.P., Bhargava D., Boltekar S.,  
BA Borokova D., Botchan M.R., Bouck J., Brockstein P., Brotler P.,  
BA Burris K.C., Busam D.A., Butler H., Cadieu L.B., Davies A., Chandra I.,  
BA Cherry J.M., Chew S.I., Dahlke C., Davidson L., Davis P.,  
BA Glodet A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
BA Harris N.L., Harvey D., Heilmann T.J., Hernandez J.R., Houck J.,  
BA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Jegham C.,  
BA Jalali M., Kalush F., Karpen G.H., Ke Z., Kemison J.A., Ketchum K.A.,  
BA Kamel B.B., Kodira C.D., Kraft C., Kravitz S., Kul D., Lai Z.,  
BA Laoko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang X., Lin X.,  
BA Liu X., Mattei B., McIntosh T.C., McLeod M.P., MCPersonson D.,  
BA Mekoulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
BA Mount S.M., Moy W., Murphy B., Murphy L., Mutny D.M., Nelson D.L.,

RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puti V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheller F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spreading A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskaas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach U.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of *Drosophila melanogaster*.";  
 RL Science 287:2185-2195(2000).

[3]  
 RP SEQUENCE FROM N.A.  
 RA Ceiniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,  
 RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,  
 RA Barzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Buesam D.A.,  
 RA Carlson J.W., Center A., Champs M., Davenport L.B., Dietz S.M.,  
 RA Carlson K., Dorsett V., Doup L.E., Doyle C., Dreonek D., Farfan D.,  
 RA Ferreira S., Frise B., Galle R.F., Gary N.S., George R.A.,  
 RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,  
 RA Ibbegam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,  
 RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nuccio J.,  
 RA Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B.,  
 RA Phanenavong S., Pittman G.S., Puti V., Richards S., Scheeler F.,  
 RA Stapleton M., Strong R., Svirskaas R., Tector C., Tyler D.,  
 RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;  
 RT "Sequencing of *Drosophila melanogaster* genome.";  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

[4]  
 RP SEQUENCE FROM N.A.  
 RA Mirra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,  
 RA Hradecky P., Huang Y., Kaminker J.S., Prochuk S.B., Smith C.D.,  
 RA Tupy J.L., Bergman C., Berman B., Carlson J.W., Ceiniker S.E.,  
 RA Clump M., Drysdale R., Emmert D., Frise B., de Grey A., Harris N.,  
 RA Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,  
 RA Searle S.M.J., Smith B., Shu S., Smurniak F., Whitfield E.,  
 RA Ashburner M., Gelbart W.M., Rubin G.M., Mangall C.J., Lewis S.E.;  
 RT "Annotation of *Drosophila melanogaster* genome.";  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

[5]  
 RP SEQUENCE FROM N.A.  
 RA Adams M.D., Ceiniker S.E., Gibbs R.A., Rubin G.M., Venter J.C.;  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

[6]  
 RP SEQUENCE FROM N.A.  
 RA FlyBase;  
 RL Submitted (SSP-2002) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF289493; AAC40806.1; -;  
 DR EMBL; AB003808; AAM70947.1; -;  
 DR FlyBase; FBgn0034075; Aeph.  
 DR InterPro; IPR001440; TPR.  
 DR InterPro; IPR006025; Zn\_MTPeptide.  
 DR PROSITE; PS00142; ZINC\_PROTEASE; 1.  
 SQ SEQUENCE 785 AA; 8963 MM; 30A8DFCD636F7F1 CRC64;

Query Match 12.0%; Score 101.5; DB 5; Length 785;  
 Best Local Similarity 24.5%; Pred. No. 21;  
 Matches 34; Conservative 29; Mismatches 51; Indels 25; Gaps 5;

QY 41 EDPLPYKGLKGYOPDGM-----EISGPEKGDAGYV-----NLSDTITK 85  
 DB 78 EDLPTPLSBSRFSK--VFDGWDBRHDDHGDHVDYDGPGEALDDHDDHDDHDDDEE 135  
 QY 86 PVFKKIRBKKEENKPT-----PVSKKKNPQVNSQLNSHKKDLQREHSQKSDS 139  
 DB 136 PLTELEBLEEBEPEPEDEPADEEYEDDEENNA--GENTTADEEBEEDND 193  
 QY 140 TKDVTATVLDKNISSKST 158  
 DB 194 EGTVEATVEATTEATTEAT 212

RESULT 31  
 ID Q9U4U6 PRELIMINARY; PRT; 948 AA.  
 AC Q9U4U6;  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-OCT-2002 (TREMBLrel. 22, Last annotation update)  
 DE Ornithine decarboxylase.  
 GN ODC.  
 OS Plasmodium falciparum.  
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
 OC NCBI\_Taxid=5833;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=FCBPL/RSa.  
 RA Birholtz L., Jobert F., Neitz A.W.H., Louw A.I.;  
 RT "Molecular characterisation of Plasmodium falciparum ornithine  
 decarboxylase cDNA obtained by RAGE.";  
 RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF139900; AAF14518.1; -;  
 DR HSP; P07805; 1F3T.  
 DR InterPro; IPR000183; Decarboxylase2.  
 DR Pfam; PF00278; Orn\_DAP\_Arg\_dec; 1.  
 DR Pfam; PF00278; Orn\_DAP\_Arg\_dec; 1.  
 DR PRINTS; PR01179; ODAACRXLASE.  
 SQ SEQUENCE 948 AA; 110350 MM; 43F103D883F12835 CRC64;

Query Match 12.0%; Score 101.5; DB 5; Length 948;  
 Best Local Similarity 21.8%; Pred. No. 26;  
 Matches 51; Conservative 31; Mismatches 63; Indels 89; Gaps 11;

QY 6 FLNKDTEVSESL-----KPRVVTYIONGKEMSTIVSEDFILPV-----KG 50  
 DB 559 YIINAGGYPBELEYNDAKGDKNHYCTLSLOEIKDKQKFLNEETFLTKGYISFEKI 618  
 QY 51 ELEKSYOPDGEISGFEKKD-----AGVINLSKDTFLKPYFKIEKKEENKPT 102  
 DB 619 SLATKMSIDHY--FSMKDKLRYICBGRRTVAASSTLVAKIIGKR-----RT 665  
 QY 103 F-----DVSKKKNPQVNSQLNSHKKDLQREHSQK--- 136  
 DB 666 FQGIWLKDLKAYDPLNFAQGENKQDEPKIKHNNDNDNNDNNNNNNNNNNKGGQ 725  
 QY 137 -----SDST-----KDYATVLDK--NNIS--SKSTTNPN 163  
 DB 726 GNIMWDLIITSTNDSTNKNKNDHSSQVIGNSCTIRDEGDNIKINTHTNPN 779

RESULT 32  
 ID Q815S6 PRELIMINARY; PRT; 1377 AA.  
 AC Q815S6;  
 DT 01-MAR-2003 (TREMBLrel. 23, Created)  
 DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)  
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)  
 DE Bacterial translation initiation factor 3 subunit 10, putative.  
 GN PF0625C.  
 OS Plasmodium falciparum (isolate 3D7).  
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
 OC NCBI\_Taxid=36329;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=3D7;  
 RX MEDLINE=22255705; PubMed=12368864;  
 RA Gardner M.J., Hall N., Fung B., White O., Berriman M., Hyman R.W.,  
 RA Carlton J.M., Pain A., Nelson K.B., Bowman S., Paulsen I.T., James K.,  
 RA Biese J.A., Nene V., Shallow S.J., Suh B., Peterson J., Angiuoli S.,  
 RA Pereira M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B.,  
 RA Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,  
 RA McPadden G.I., Cummings L.M., Subramanian G.M., Mangall C.,  
 Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,

RA Pfaffar C.M., Barrell B.;  
RT "Genome sequence of the human malaria parasite *Plasmodium*  
falciparum";  
RL Nature 419:498-511(2002).  
DR EMBL; AF014846; AAN36214.1; -  
KW Initiation factor.  
SQ SEQUENCE 1377 AA; 166059 MW; 587CF2B3FC8FBEB9 CRC64;

Query Match 12.0%; Score 101.5; DB 5; Length 1377;  
Best Local Similarity 22.3%; Pred. No. 39;  
Matches 43; Conservative 37; Mismatches 62; Indels 51; Gaps 9;

QY 3 VKPFLNDTGVSELSKRVTVTLONGKE--NSSTIVSEBDFILPYVKGLEKGYQF-- 58  
DB 619 IKDSLNNYNDLQNGELKMINETIALENEVENITMSTLTTEEEF---AYLDKSNKMINNT 675  
QY 59 ---DGMETSGPEGKXGAGYVNLKDPYFKPFIKIEK-----KEENKPTFDVSK 107  
DB 676 ATLBDIINVEFEKRP-----INNEKLLKITYEKIDEBHKKIQLSEHNKKKELIX 728  
QY 108 KKDNPVNSQLN-----ESHKEDLQR-----EHSQKSDSTQVAT 146  
DB 729 KQ--KEFGAQKLMKEKLEKLEKLEKLEKLEKLEKLEKLEKLEKLEKLEKLEKLEK 786  
QY 147 VLDKNNISSKST 159  
DB 787 I---KKLCSTNTT 796

RESULT 33  
081353 PRELIMINARY; PRT; 1373 AA.  
AC 081353;  
DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
DE Helicase with Zn-finger motif, putative.  
GN PF10480W.  
OS Plasmodium falciparum (isolate 3D7).  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
OC NCBI\_TaxID=36329;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=22255708; PubMed=12368867;  
RA Hall N., Pain A., Bettiman M., Chucher C., Harris B., Harris D.,  
Munhall K., Bowman S., Atkin R., Baker S., Barton A., Brooks K.,  
Buckee C.O., Burrows C., Cherevach I., Chillingworth C.,  
Chillingworth T., Christodoulou Z., Clark L., Clark R., Corton C.,  
Cronin A., Davies R., Davis P., Dear P., Dearden F., Doggett J.,  
Fellwell T., Goble A., Goodhead I., Gwilliam R., Hamlin N., Hance Z.,  
Harper D., Hauser H., Hornsby T., Holroyd S., Horrocks P.,  
Humphrey S., Jagels K., James K.D., Johnson D., Kerhroux A.,  
Knights A., Konfortov B., Kyes S., Latke N., Lawson D., Lennard N.,  
Line A., Maddison M., Mclean J., Mooney P., Moule S., Murphy L.,  
Oliver K., Ormond D., Price C., Quail M.A., Rabinowitsch E.,  
Rajandream M.A., Rutter S., Rutherford K.M., Sanders M., Simmonds M.,  
Seeger K., Sharp S., Smith R., Squares R., Squares S., Stevens K.,  
Taylor K., Tivey A., Unwin L., Whitehead S., Woodward J.,  
Sutcliffe J.E., Craig A., Newbold C., Barrell B.G.  
RT "Sequence of *Plasmodium falciparum* chromosomes 1, 3-9 and 13.";  
RL Nature 419:527-531(2002).  
DR EMBL; AL929356; CAD51782.1; -  
SQ SEQUENCE 1373 AA; 160372 MW; 059F84B99EBF4702 CRC64;

Query Match 11.8%; Score 100.5; DB 5; Length 1373;  
Best Local Similarity 21.7%; Pred. No. 45;  
Matches 40; Conservative 33; Mismatches 54; Indels 57; Gaps 9;

QY 6 FLINDTGVSE---LKQHRVTVTLONGKEMSGSTIVSEDFILPYVKGLEKGYQF-- 61  
DB 418 YVNDDEGRVIGWESITMLPHRVOIL-----LSATV-----PNT-----LEFADW 457  
QY 62 -----EISGFEKK-----DAGVYINLSKDTPIKPVK-----KIEKKKEE 98

DB 458 VGFTRKQKAVISISTKRPVPLHYIYVDSYVLVNDKKNYSSAFKELIYKIREKQZAN 517  
QY 99 NKPTFDVSKKDNPOVNSQLNESH-RKEDLQREHSQKSDSTQVATVLDKNNISSKS 157  
DB 518 NNTKQITSGSNSSNLKNNNYDSKXKILTTNNKENDVT-----QNNNNNNN 568  
QY 158 TTN 161  
DB 569 NNNN 572

RESULT 34  
014712 PRELIMINARY; PRT; 375 AA.  
AC 014712;  
DT 01-JAN-1998 (TrEMBLrel. 05, Created)  
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
DE Cell cycle progression restoration 8 protein.  
GN CPR8.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OC NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=98043401; PubMed=9383053;  
RA Edwards M.C., Liegeois N., Horecka J., Depinho R.A., Sprague G.F. Jr.,  
Tyers M., Bledge S.J.,  
RT "Human CPR (cell cycle progression restoration) genes impart a Far-  
phenotype on yeast cells".  
RL Genetics 147:1063-1076(1997).  
DR EMBL; AF011794; AAB69314.1; -  
DR InterPro: IPR004238; LEA.  
DR Pfam: PF02987; LEA; 1.  
SQ SEQUENCE 375 AA; 44320 MW; 0C261BF68AAEF7AD CRC64;

Query Match 11.8%; Score 100; DB 4; Length 375;  
Best Local Similarity 28.0%; Pred. No. 12;  
Matches 51; Conservative 16; Mismatches 67; Indels 48; Gaps 9;

QY 13 GEVSELKPRVTVTLONGKEMSGSTIVSE--EDFILPYVKGLEKGYQF-----DGMET--- 62  
DB 18 GELQQLSSQL-----HGKSDSPNYTEKELAILNERLTLEKLTFCQSSDLWBRLY 72  
QY 63 -----ISGFEKKDAG-----YVINSKDTPIKVPFIKIEKKEENKPTFDVSKKD 110  
DB 73 VEAKDNGKQGTGDKKKGGSGSHRVNKKSGTFLGVS-----KETFDAMKNS 120  
QY 111 NQVNSQLNESHKEDLQREHSQKSDST-----KQVATVLD-KNNISSKSTTNP 162  
DB 121 KEFVRHHEKIKQAKDV-KENLKPSDSVSKTFPHKQTTQNIIFDEKKNKRNATKAA 179  
QY 163 NK 164  
DB 180 EK 181

RESULT 35  
0980A1 PRELIMINARY; PRT; 622 AA.  
AC 0980A1;  
DT 01-OCT-2001 (TrEMBLrel. 18, Created)  
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)  
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)  
DE LIPOPROTEIN.  
GN MYPU\_4650.  
OS Mycoplasma pulmonis.  
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmales; Mycoplasma.  
OC NCBI\_TaxID=2107;  
RN [1]  
RP SEQUENCE FROM N.A.

RC STRAIN=UAB CTIP;  
 RX MEDLINE=21267165; PubMed=11353084;  
 RA Chambud I., Heilig R., Ferris S., Barbe V., Samson D., Galisson F.,  
 RA Moszer I., Dybvig K., Wroblewski H., Viari A., Rocha B.P.C.,  
 RA Blanchard A.;  
 RT "The complete genome sequence of the murine respiratory pathogen  
 RL Mycoplasma pulmonis";  
 RU Nucleic Acids Res. 29;2145-2153(2001).  
 DR EMBL; AL445564; CAC13638.1; -.  
 DR MyPulist; MYPUL\_4650; -.  
 KM Complete proteome.  
 SQ SEQUENCE 622 AA; 73762 MW; 42BD88930861960D CRC64;

Query Match 11.8%; Score 100; DB 16; Length 622;  
 Best Local Similarity 24.4%; Pred. No. 21;  
 Matches 40; Conservative 31; Mismatches 55; Indels 38; Gaps 7;

QY 8 LNKDGEVSELEKPHRVVTYTIQNGKMSSTVSEEDFILPVYKGELEKGYOFGWEISGPE 67  
 DB 69 INSETEIVKKE-----TILN-----LSEDDITFSLNTENNAFQULDEF----- 108  
 QY 68 GKXDAGYVINSKDTPIKVFPEKIEKKEEKNKPPDVSKKD-----NPVNHSQLNES- 122  
 DB 109 -----VSKDEKFKIKFOEINFSQTEQ-KITDNISKEDEKKNPKDNENNNSS 157  
 QY 123 -HRKDELQREHRSQKSDSTKQDVTATVLDKN-NISSKSTNNPNK 164  
 DB 158 DQKDELQKNSDKLNDVQDEKANKENSNSDSKENDENTNK 201

## RESULT 36

081J24 PRELIMINARY; PRT; 1130 AA.

ID 081J24;  
 AC 081J24;  
 DT 01-MAR-2003 (TREMBLrel. 23, Created)  
 DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)  
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)  
 DE Hypothetical protein.  
 GN PF10\_0046.  
 OS Plasmodium falciparum (isolate 3D7).  
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
 NC NCB1\_TaxID=36329;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=3D7;  
 RX MEDLINE=22255705; PubMed=12368864;  
 RA Gardner M.J., Hall N., Fung E., White O., Bertman M., Hyman R.W.,  
 RA Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,  
 RA Bisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,  
 RA Chan M.-S., Nene V., Shallow S.J., Sub B., Peterson J., Angiuoli S.,  
 RA Pereira M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B.,  
 RA Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,  
 RA McCadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,  
 RA Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.M.,  
 RA Fraser C.M., Barrell B.;  
 RT "Genome sequence of the human malaria parasite Plasmodium  
 RL falciparum";  
 RU Nature 419;498-511(2002).  
 DR EMBL; AB014829; AAN35244.1; -.  
 KM Hypothetical protein.  
 SQ SEQUENCE 1130 AA; 131697 MW; FB6AAFB08C4CCDB6 CRC64;

Query Match 11.8%; Score 100; DB 5; Length 1130;  
 Best Local Similarity 26.1%; Pred. No. 40;  
 Matches 30; Conservative 18; Mismatches 53; Indels 14; Gaps 3;

QY 64 SGFEKGDAGYV-INLSKDTPIKVFPEKIEKKEEKNKPPDVSKKDNPVNHSQLNE 121  
 DB 568 SFLSGNSDNYRIDRIYLSPDNYFYSNNRNINDENEVYQVETRSADYNSHDETNE 627  
 QY 122 SHRKDELQREHRSQKSDSTKQDVTATVLDKN-NISSKSTNNPNK 164  
 DB 628 NNEKEINNEBGNBEBKSKRKKTNDYDNDKQVYNNASGNSNHSNHNNNNNK 682

## RESULT 37

08IDA0 PRELIMINARY; PRT; 3127 AA.

ID 08IDA0;  
 AC 08IDA0;  
 DT 01-MAR-2003 (TREMBLrel. 23, Created)  
 DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)  
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)  
 DE Guanylyl cyclase (BC 4.6.1.2).  
 GN GC-BETA.  
 OS Plasmodium falciparum (isolate 3D7).  
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
 NC NCB1\_TaxID=36329;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RA Harris B., Lemnard N., Clark L., line A., Barron A., Corron C.,  
 RA Berriman M., Pain A., Hall N., Atkin R., Chillingworth C., Doggett J.,  
 RA Ormond D., Sanders M., Hayes R., Hall S., Quail M., Barrell B.;  
 RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AL844509; CAD52725.1; -.  
 KM Lyase.  
 SQ SEQUENCE 3127 AA; 370597 MW; F0375C72B9007560 CRC64;

Query Match 11.8%; Score 100; DB 5; Length 3127;  
 Best Local Similarity 24.2%; Pred. No. 1.2e+02;  
 Matches 52; Conservative 34; Mismatches 71; Indels 58; Gaps 9;

QY 4 KEFLNKDGEVSELEKPHRVVTYTIQNG-----KEMSSVSEEDFILPVYKGELEKGYO 57  
 DB 1969 KKF--KRNYSVYLESPLHLIGDIVDNINIRKKKKKKKIKITVSDNFTSPVNIKEVYNEQ 2026  
 QY 58 FDGWEISG---FEKGDAGYVINSKDTPIKVFPEKIEKKEEKNK----- 100  
 DB 2027 ERKEIVGNLSYDKTKKIPPIKFTKEGRIRK--KKIEKKEKKEKKNNNNNFLYNDYSS 2084  
 QY 101 ---PTF-----DVSKKKNPQVNHSQL-----NESHK--EDLQR 130  
 DB 2085 YSSPKYGNENNVFKYIRERKDPQKTDHPFNFNSKFLHNYNPMKNKKNKNKNVNR 2144  
 QY 131 EHSQKSDSTKD-VTATVLDKNNISKSTNNPNK 164  
 DB 2145 NEYPVYSSSKDGYVNFSLDSLPSSDNEYSDDNE 2179

## RESULT 38

093424 PRELIMINARY; PRT; 385 AA.

ID 093424;  
 AC 093424; P90801;  
 DT 01-JAN-1999 (TREMBLrel. 09, Created)  
 DT 01-JAN-1999 (TREMBLrel. 09, Last sequence update)  
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)  
 DE Hypothetical glycine-rich 37.0 kDa protein E02A10.2 in chromosome V  
 DB precursor.  
 GN E02A10.2.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;  
 OC Rhabditidae; Peloderinae; Caenorhabditis.  
 NC NCB1\_TaxID=6239;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BRISTOL N2;  
 RA Thomas K.;  
 RL Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.  
 RN (2)  
 RP REVISIONS.  
 RC STRAIN=BRISTOL N2;  
 RA Jones S.J.M.;  
 RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; Z81053; CAB02877.1; -.  
 DR HSP; P10968; 2CMG.  
 DR WormRep; E02A10.2; CB09116.  
 DR Interpro; IPR002952; Eggshe11.

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 10, 2004, 10:48:44 ; Search time 35.0346 Seconds  
(without alignments)  
697.707 Million cell updates/sec

Title: US-10-067-385-8\_COPY\_620\_773

Perfect score: 799  
Sequence: 1 DTGEVSLKPRVTVTIQNG.....ATVLDKNNISSKSTNNPK 154

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: A\_Geneseq\_19jun03:\*  
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:\*  
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:\*  
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:\*  
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:\*  
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:\*  
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:\*  
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:\*  
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:\*  
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:\*  
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:\*  
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:\*  
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:\*  
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:\*  
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:\*  
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:\*  
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:\*  
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:\*  
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:\*  
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:\*  
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:\*  
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:\*  
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:\*  
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:\*  
25: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	799	100.0	773	22	AA848343 S. pneumoniae Sp13
2	799	100.0	2120	24	AAV81710 Streptococcus pneu
3	799	100.0	2140	24	ABU01020 S. pneumoniae type
4	615	77.0	117	19	AAW55096 Streptococcus pneu
5	615	77.0	117	23	ABP54590 S. pneumoniae SP04
6	118	14.8	746	22	AA681779 S. epidermidis ope
7	118	14.8	778	23	ABP39023 Staphylococcus epi
8	107.5	13.5	484	21	AA647777 Arabidopsis thalia
9	105	13.1	665	21	AA818278 Plasmodium falcipa

10	101.5	12.7	564	22	AB861977 Drosophila melanog
11	96.5	12.1	2060	23	AAE20967 Staphylococcus lug
12	96	12.0	2500	21	AA818272 Plasmodium falcipa
13	93.5	11.7	408	21	AA637134 Arabidopsis thalia
14	93.5	11.7	456	21	AA637133 Arabidopsis thalia
15	93.5	11.7	476	21	AA637132 Arabidopsis thalia
16	92.5	11.6	2519	22	ABG16636 Novel human diagno
17	90.5	11.3	209	21	AA842897 Human ORF2661
18	90.5	11.3	757	21	AAV44364 Human cell cycle r
19	90.5	11.3	758	21	AA853319 Human colon cancer
20	90	11.3	442	23	ABP38188 Staphylococcus epi
21	90	11.3	645	24	ABJ19106 Staphylococcus epi
22	90	11.3	654	24	ABP56879 Staphylococcus epi
23	90	11.3	1183	22	AB858769 Drosophila melanog
24	89.5	11.2	402	20	AAV03190 S. aureus trigger
25	89.5	11.2	511	20	AAV35091 Chlamydia pneumoni
26	89.5	11.2	525	20	AAV03189 S. aureus trigger
27	89.5	11.2	1408	22	AB858704 Drosophila melanog
28	89	11.1	645	24	ABJ18879 Pathogen specific
29	88.5	11.1	281	22	AB864828 Staphylococcus epi
30	88.5	11.1	1141	22	AA850008 Shrimp white spot
31	88	11.0	225	23	ABP73392 Candida albicans e
32	88	11.0	258	22	AA894584 Human protein seq
33	88	11.0	817	22	AAW79318 Human protein SEQ
34	88	11.0	817	22	AAW79319 Human protein SEQ
35	87.5	11.0	635	23	ABP73209 Candida albicans e
36	87.5	11.0	2515	22	AB812281 Human secreted pro
37	87.5	11.0	2515	22	AA880268 Human protein SEQ
38	87.5	11.0	2515	22	AA880269 Human protein SEQ
39	87.5	11.0	3021	24	ABP75873 Human secretory po
40	87	10.9	313	21	AA647057 Arabidopsis thalia
41	87	10.9	607	22	AAW39950 Human polypeptide
42	87	10.9	635	21	AA818176 Plasmodium falcipa
43	87	10.9	902	23	ABP40312 Staphylococcus epi
44	87	10.9	922	22	AA850023 Shrimp white spot
45	86.5	10.8	443	22	AAU38439 Salmomella typhi c

## ALIGNMENTS

RESULT 1	AA848343
ID	AA848343 standard; Protein; 773 AA.
XX	AA848343;
XX	20-APR-2001 (first entry)
DT	S. pneumoniae Sp130 polypeptide.
XX	
DE	Immunogenic; Sp128; Sp130; pneumococcal; otitis media; nasopharyngeal;
KW	bronchial; lung; blood; infection; immune response; immunotherapy;
KW	antibacterial; auditory; vaccine.
XX	
OS	Streptococcus pneumoniae.
XX	
PN	MO200076540-A2.
XX	
PD	21-DEC-2000.
XX	
PP	09-JUN-2000; 2000MO-US15925.
XX	
PR	10-JUN-1999; 99US-0138453.
XX	
PA	(MED1-) MED IMMUNE INC.
XX	
PI	Adamou JE, Choi GH;
XX	
XX	WPI; 2001-112197/L2.
DR	N-PSDB; AAC84742.
XX	
PT	New vaccines comprising Sp128 or Sp130 polypeptides, for treating and

PT preventing pneumococcal infections, particularly infections caused by  
PT Streptococcus, e.g. otitis media, nasopharyngeal, bronchial, lung or  
PT blood infections -  
XX  
XX  
PS Claim 8; Page 51-54; 54pp; English.  
XX  
XX The invention relates to novel immunogenic polypeptides, Sp128 and Sp130  
CC from *S. pneumoniae*. Vaccines comprising the polypeptides are useful for  
CC the treatment and prevention of pneumococcal infections, particularly  
CC infections caused by *Streptococcus*, such as otitis media, nasopharyngeal,  
CC bronchial, lung or blood infections. The antigens are used as immunogenic  
CC agents to stimulate an immune response. The antibodies and antibodies may  
CC also be used in diagnosing and treating pneumococcal infections.  
CC Recombinant polypeptides serve as a mechanism for stimulating production  
CC of antibodies for use in passive immunotherapy, diagnostic reagents, and  
CC as reagents in other processes such as affinity chromatography. The  
CC present sequence represents the *S. pneumoniae* Sp130 polypeptide.  
XX  
XX  
SQ Sequence 773 AA;  
Query Match 100.0%; Score 799; DB 22; Length 773;  
Best Local Similarity 100.0%; Pred. No. 2.3e-71;  
Matches 154; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 DTGEVSELKPHRYVTVTIÖNGKEMSSSTIVSEDFILPYKGLKGYQFDGWEISGFEGKK 60  
DB 620 DTGEVSELKPHRYVTVTIÖNGKEMSSSTIVSEDFILPYKGLKGYQFDGWEISGFEGKK 679  
QY 61 DAGVINLSKDTPIKPYFKKIEKKKEENKPTFVSKKKKNPQVNHSQLNESHKEDLÖR 120  
DB 680 DAGVINLSKDTPIKPYFKKIEKKKEENKPTFVSKKKKNPQVNHSQLNESHKEDLÖR 739  
QY 121 EESHÖKS DSTKDVTATVLDKNNISSKSTNNPNK 154  
DB 740 EESHÖKS DSTKDVTATVLDKNNISSKSTNNPNK 773  
RESULT 2  
AAV81710  
ID AAV81710 standard; Protein; 2120 AA.  
XX  
XX AAV81710;  
AC  
XX  
DT 02-JUN-2000 (first entry)  
XX  
XX Streptococcus pneumoniae protein sequence ID3.  
DE  
XX Streptococcus pneumoniae infection; immunogen; antigen; diagnosis; AIDS;  
KW bacterial pneumonia; asplenia; heart disease; lung disease; alcoholism;  
KW kidney disease; diabetes; immunosuppressive disorder; otitis media;  
KW pneumococcal septicemia; sinusitis; meningitis; therapy.  
XX  
XX Streptococcus pneumoniae.  
OS  
XX  
XX WO200006738-A2.  
PN  
XX  
XX 10-FEB-2000.  
PD  
XX  
XX 27-JUL-1999; 99MO-GB02452.  
PP  
XX  
XX 27-JUL-1998; 98GB-0016336.  
PR  
XX 19-MAR-1999; 99US-0125329.  
PR  
XX  
XX (MICR-) MICROBIAL TECHNICS LTD.  
PA  
XX  
XX Le Page RWF, Wells JM, Hannify SB, Hansbro PM;  
PI  
XX  
XX WPI; 2000-195301/17.  
DR  
XX N-PSDB; AAZ91806.  
DR  
XX  
XX Streptococcal proteins and polynucleotides useful for diagnosis,  
PT treatment and prophylaxis of bacterial infections -  
XX

PS Claim 2; Page 41-42; 76pp; English.  
XX  
XX This sequence represents a Streptococcus pneumoniae protein of the  
CC invention. The proteins (or their homologues, derivatives and/or  
CC fragments) are useful as immunogens or antigens. Immunogenic or antigenic  
CC compositions comprising the proteins are useful as vaccines and also in  
CC diagnostic assays. The sequences are useful for the detection or  
CC diagnosis of *S. pneumoniae* infection, by contacting a sample to be tested  
CC with them. Agents capable of antagonising, inhibiting or interfering with  
CC the function or expression of the protein or polypeptide are useful in  
CC medical compositions in the treatment or prophylaxis of *S. pneumoniae*  
CC infection. As the sequences can be used to treat *S. pneumoniae* infection,  
CC they can be used to treat bacterial pneumonia, which has high rates in  
CC young children, the elderly, and in patients with predisposing conditions  
CC such as asplenia, heart, lung and kidney disease, diabetes, alcoholism,  
CC or with immunosuppressive disorders, especially AIDS. They can also be  
CC used to treat pneumococcal septicemia, otitis media, sinusitis, and  
CC meningitis.  
XX  
XX  
SQ Sequence 2120 AA;  
Query Match 100.0%; Score 799; DB 21; Length 2120;  
Best Local Similarity 100.0%; Pred. No. 9.3e-71;  
Matches 154; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 DTGEVSELKPHRYVTVTIÖNGKEMSSSTIVSEDFILPYKGLKGYQFDGWEISGFEGKK 60  
DB 1933 DTGEVSELKPHRYVTVTIÖNGKEMSSSTIVSEDFILPYKGLKGYQFDGWEISGFEGKK 1992  
QY 61 DAGVINLSKDTPIKPYFKKIEKKKEENKPTFVSKKKKNPQVNHSQLNESHKEDLÖR 120  
DB 1993 DAGVINLSKDTPIKPYFKKIEKKKEENKPTFVSKKKKNPQVNHSQLNESHKEDLÖR 2052  
QY 121 EESHÖKS DSTKDVTATVLDKNNISSKSTNNPNK 154  
DB 2053 EESHÖKS DSTKDVTATVLDKNNISSKSTNNPNK 2086  
RESULT 3  
ABU01020  
ID ABU01020 standard; Protein; 2140 AA.  
XX  
XX ABU01020;  
AC  
XX  
DT 11-FEB-2003 (first entry)  
XX  
XX  
DE S. pneumoniae type 4 strain protein from coding region #590.  
XX  
XX Bacterial meningitis; pneumonia; sepsis; otitis media;  
KW ear infection; antiinflammatory; antibacterial; immunostimulant;  
KW auditory; respiratory; gene therapy; vaccine.  
XX  
XX Streptococcus pneumoniae type 4 strain.  
OS  
XX  
XX WO200277021-A2.  
PN  
XX  
XX 03-OCT-2002.  
PD  
XX  
XX 27-MAR-2002; 2002MO-IB02163.  
PP  
XX  
XX 27-MAR-2001; 2001GB-0007658.  
PR  
XX  
XX (CHIR-) CHIRON SPA.  
PA  
XX (GENO-) INST GENOMIC RES.  
PA  
XX  
XX Masignani V, Tettelin H, Fraser C;  
PI  
XX  
XX WPI; 2003-040579/03.  
DR  
XX N-PSDB; ABX06302.  
DR  
XX  
XX New proteins and nucleic acid molecules from Streptococcus pneumoniae,  
PT useful as medicaments for treating or preventing a disease or infection  
PT due to streptococcus bacteria, such as pneumonia, sepsis, otitis media



PT or ear infection -  
XX  
PS Claim 1; SEQ ID No 1180; 56pp; English.  
XX  
XX The invention relates to a protein comprising or having at least 50%  
CC identity to any of the 2469 amino acid sequences, identified in the  
CC specification (available on a computer readable format), or its fragment,  
CC expressed from 2469 of 2489 identified DNA coding regions from the  
CC Streptococcus pneumoniae type 4 strain genomic sequence appearing as  
CC AB556454. Also included are an antibody which binds one of the  
CC proteins, treating a patient by administering the protein, DNA or  
CC antibody (in a composition), a kit comprising first and second primers,  
CC which are the nucleic acid cited above or fragments between nucleotides  
CC 8-100 of a sequence not defined in the specification, for amplifying a  
CC target sequence contained within a Streptococcus nucleic acid sequence,  
CC where the first primer is substantially complementary to the target  
CC sequence and the second primer is substantially complementary to the  
CC complement of the target sequence, and where the parts of the primers  
CC having substantial complementarity define the termini of the target  
CC sequence to be amplified, assay comprising contacting a test compound  
CC with the protein, and determining whether the test compound binds to the  
CC protein and a Streptococcus pneumoniae bacterium, where one or more  
CC genes encoding the proteins has been rendered inactive. The proteins,  
CC nucleic acid molecules, antibody and compositions are useful as  
CC medicaments for treating or preventing a disease or infection due to  
CC streptococcus bacteria, particularly *S. pneumoniae*, such as pneumonia,  
CC sepsis, otitis media or ear infection. They are also useful in developing  
CC vaccines, diagnostics and antibiotics. The methods are useful for  
CC identifying immunodominant proteins. The present sequence is one of  
CC the 2469 proteins expressed by the identified coding regions from the  
CC genomic sequence.  
CC Note: The sequence data for this patent did not form part  
CC of the printed specification, but was obtained in electronic  
CC format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 2140 AA;  
Query Match 100.0%; Score 799; DB 24; Length 2140;  
Best Local Similarity 100.0%; Pred. No. 9,4e-71;  
Matches 154; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 DTGVSALKRRVVTITONGKMSSTVSEDFILPVYKGLKGYQPDGWEISGFEKK 60  
DB 1953 DTGVSALKRRVVTITONGKMSSTVSEDFILPVYKGLKGYQPDGWEISGFEKK 2012  
QY 61 DAGVITLSTKPTFKPVPKKEEKEENKPTFVSKKKQNPQVNHSLNSHKKEDL 120  
DB 2013 DAGVITLSTKPTFKPVPKKEEKEENKPTFVSKKKQNPQVNHSLNSHKKEDL 2072  
QY 121 EHSQKSDSTKDVATVLDKNNISSKSTNNPK 154  
DB 2073 EHSQKSDSTKDVATVLDKNNISSKSTNNPK 2106  
RESULT 4  
ID AAM55096 standard; Protein; 117 AA.  
XX  
AC AAM55096;  
XX  
DT 02-OCT-1998 (first entry)  
XX  
DE Streptococcus pneumoniae SP0043 protein.  
XX  
KW Streptococcus pneumoniae; antigen; vaccine; infection; diagnosis;  
KW detection; pneumonia; otitis media; meningitis.  
XX  
OS Streptococcus pneumoniae.  
XX  
PN W09818930-A2.  
XX  
PD 07-MAY-1998.

XX  
PP 30-OCT-1997; 97MO-US19422.  
XX  
XX  
PR 31-OCT-1996; 96US-0029960.  
XX  
XX (HUMA-) HUMAN GENOME SCI INC.  
PA  
PI Choi GH, Hromockyj A, Johnson LS, Kunsch CA;  
XX  
XX WPI; 1998-272224/24.  
DR  
DR N-PSDB; AAV27357.  
XX  
PT Nucleic acid encoding antigenic peptide(s) from Streptococcus  
PT pneumoniae - or their epitope-containing fragments, useful in  
PT protective or therapeutic vaccines, and for diagnosis  
XX  
PS Claim 11; Page 62; 118pp; English.  
XX  
XX The present sequence represents a protein from Streptococcus pneumoniae.  
CC The nucleic acid sequence encoding the Streptococcus pneumoniae protein  
CC can be useful in vaccines for inducing protective antibodies against  
CC Streptococcus pneumoniae, for treatment or prevention of infection e.g.  
CC pneumonia, otitis media or meningitis. Probes based on the nucleic acid  
CC are used to detect Streptococcus infection (by usual hybridisation or  
CC amplification methods), also for isolating Streptococcus genes or their  
CC allelic variants. The protein can be used similarly to detect specific  
CC antibodies in standard immunoassays, especially for diagnosing or  
CC monitoring infections. Antibodies which bind the protein are used to  
CC detect corresponding antigens, to purify the protein and for passive  
CC immunisation (optionally coupled to a toxin). Vaccines are administered,  
CC e.g. by injection, orally or through the skin, typically at 0.01-1000  
CC (especially 10-300) mu g/ml per dose.  
XX  
SQ Sequence 117 AA;  
Query Match 77.0%; Score 615; DB 19; Length 117;  
Best Local Similarity 100.0%; Pred. No. 5,2e-54;  
Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 38 YKGELEKGYQPDGWEISGFBGKQAGVYINLSKDTFIPVVKKEEKEENKPTFVSK 97  
DB 1 YKGELEKGYQPDGWEISGFBGKQAGVYINLSKDTFIPVVKKEEKEENKPTFVSK 60  
QY 98 KQNPQVNHSLNSHKKEDLQREHSHQKSDSTKDVATVLDKNNISSKSTNNPK 154  
DB 61 KQNPQVNHSLNSHKKEDLQREHSHQKSDSTKDVATVLDKNNISSKSTNNPK 117  
RESULT 5  
ID ABP54590 standard; Protein; 117 AA.  
XX  
AC ABP54590;  
XX  
DT 04-SEP-2002 (first entry)  
XX  
DE S. pneumoniae SP043 protein sequence SEQ ID NO:68.  
XX  
KW Streptococcus pneumoniae; epitope; vaccine; antigenic protein;  
KW antibacterial; Streptococcal infection; detection.  
XX  
OS Streptococcus pneumoniae.  
XX  
PN US2002061545-A1.  
XX  
PD 23-MAY-2002.  
XX  
PP 22-JAN-2001; 2001US-0765272.  
XX  
PR 30-OCT-1997; 97US-0961083.  
XX  
PA (CHOI/) CHOI G H.  
PA (KUNSCH/) KUNSCH C A.

PA (BARA/) BARASH S C.  
 PA (DILL/) DILLON P J.  
 PA (DOUG/) DOUGHERTY B.  
 PA (FANN/) FANNON M R.  
 PA (ROSE/) ROSEN C A.  
 XX  
 PI Choi GH, Kunsch CA, Barash SC, Dillon PJ, Dougherty B, Fannon MR;  
 PI Rosen CA;  
 XX  
 DR MPI; 2002-479261/51.  
 DR N-PSDB; ABQ84825.  
 XX  
 PT New Streptococcus pneumoniae antigens, useful for detecting  
 PT Streptococcus and for preventing or attenuating disease caused by  
 PT Streptococcus infection -  
 XX  
 PS Claim 11; Page 29; 70pp; English.  
 XX  
 CC ABQ84792 to ABQ84904 represents nucleic acids which encode the  
 CC Streptococcus pneumoniae antigens given in ABP54557 to ABP54669.  
 CC The S. pneumoniae antigens have antibacterial activity and can be  
 CC used in vaccines. The S. pneumoniae antigens can also be used to  
 CC prevent or attenuate a Streptococcal infection in an animal. The  
 CC polynucleotides encoding the S. pneumoniae antigens can be used to  
 CC detect Streptococcus nucleic acids. ABQ84905 to ABQ85130 represent  
 CC primers used in the cloning of S. pneumoniae ORFs (open reading frames)  
 CC which are used in an example from the present invention.  
 CC  
 SQ Sequence 117 AA;  
 XX  
 Query Match 77.0%; Score 615; DB 23; Length 117;  
 Best Local Similarity 100.0%; Pred. No. 5.2e-54;  
 Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 36 YKGELEKGYQPDGWEISGFEGKKDAGYVTLNLSKQTFIKPVFKIEKKKEENKPTFDVSK 97  
 DB 1 YKGBLEKGYQPDGWEISGFEGKKDAGYVTLNLSKQTFIKPVFKIEKKKEENKPTFDVSK 60  
 QY 98 KKNPQVNHSQLNESHKEDLQREHSHQKSDSTKDVATVLDKNNISSKSTNNPK 154  
 DB 61 KKNPQVNHSQLNESHKEDLQREHSHQKSDSTKDVATVLDKNNISSKSTNNPK 117  
 XX  
 RESULT 6  
 AAG81779  
 ID AAG81779 standard; Protein; 746 AA.  
 XX  
 AC AAG81779;  
 XX  
 DT 03-SEP-2001 (first entry)  
 XX  
 DB S. epidermidis open reading frame protein sequence SEQ ID NO:652.  
 XX  
 KW Staphylococcus epidermidis SRI strain; infection; diagnosis;  
 XX  
 KM vaccination; endocarditis.  
 XX  
 OS Staphylococcus epidermidis.  
 XX  
 PN WO200134809-A2.  
 XX  
 PD 17-MAY-2001.  
 XX  
 PF 09-NOV-2000; 2000WO-US30782.  
 XX  
 PR 09-NOV-1999; 99US-0164258.  
 XX  
 PA (GLAX ) GLAXO GROUP LTD.  
 XX  
 PI Kimmery WJ;  
 XX  
 DR MPI; 2001-316495/33.  
 DR N-PSDB; AAH52629.  
 XX

PT Nucleic acids encoding polypeptides from Staphylococcus epidermidis.  
 PT Useful for vaccinating against infections, e.g. endocarditis -  
 XX  
 PS Claim 18; Page 208; 218pp; English.  
 XX  
 CC AAH52304 to AAH53970 represent nucleic acids (i) encoding polypeptides  
 CC (ii), given in AAG81454 to AAG83320, from Staphylococcus epidermidis.  
 CC (i) and (ii) can have antibacterial activity and therefore can be used  
 CC in vaccination. The nucleic acids (i) may be used to produce the  
 CC S. epidermidis polypeptides (ii) via the production of vectors  
 CC containing them which are used to produce hosts cells which express the  
 CC polypeptides. The polypeptides (ii) (and/or nucleic acids) may then be  
 CC used to vaccinate subjects and to raise antibodies against the bacteria.  
 CC The polypeptides may also be used to assay for other inhibitors of their  
 CC activity and therefore identify compounds that may be used for the  
 CC treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to  
 CC AAH5090 represent specifically claimed S. epidermidis genome DNA  
 CC polynucleotide sequences from the present invention. AAH5091 to  
 CC AAH5098 represent oligonucleotide sequences and primers which are used  
 CC in the exemplification of the present invention.  
 CC N.B. The present invention specifically claims all the polynucleotide  
 CC sequences given in the sequence listing of the present specification,  
 CC however the sequence listing only goes up to SEQ ID NO:4454 so even  
 CC though sequences are given in the disclosure for SEQ ID NO:4465 to 4472,  
 CC no sequences are present for SEQ ID NO:4455 to 4464.  
 CC  
 SQ Sequence 746 AA;  
 XX  
 Query Match 14.8%; Score 118; DB 22; Length 746;  
 Best Local Similarity 27.2%; Pred. No. 0.0063;  
 Matches 47; Conservative 23; Mismatches 57; Indels 46; Gaps 9;  
 QY 1 DTGESELEKHPHVTYTIQNGKMSSTIYSEEDFILPVYK-----GELEKGYQPDGWEIS 51  
 DB 586 DSVNAGSLKP-----ITIGKQKIQKQSVKSGTKVLPKSKVMTLMTDGLTMP-DNYGTWKE 640  
 QY 52 EISGFE-----GKKDAGYVTLNLSKQTFIKPVFKIEKKKEENKPTFDVSK 98  
 DB 641 DVLAFEDLTKLVSTKNGFVFNOSISKQITK-----NKKIETVLSAEDT 687  
 QY 99 KKNPQVNHSQLNESHKEDLQREHSHQKSDSTKDVATVLDKNNISSKSTNNPK 151  
 DB 688 DDDQKTDSDSSDKKDKADKEDHSNTSSSTKN-----DKSNADS KDDSD 724  
 XX  
 RESULT 7  
 ABP39023  
 ID ABP39023 standard; Protein; 778 AA.  
 XX  
 AC ABP39023;  
 XX  
 DT 24-UTL-2002 (first entry)  
 XX  
 DB Staphylococcus epidermidis ORF amino acid sequence SEQ ID NO:3868.  
 XX  
 KW Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;  
 XX  
 KM antibacterial; gene therapy.  
 XX  
 OS Staphylococcus epidermidis.  
 XX  
 PN US6380370-B1.  
 XX  
 PD 30-APR-2002.  
 XX  
 PF 13-AUG-1998; 98US-0134001.  
 XX  
 PR 14-AUG-1997; 97US-055779P.  
 PR 08-NOV-1997; 97US-064964P.  
 XX  
 PA (GENO-) GENOME THERAPEUTICS CORP.  
 XX  
 PI Doucette-Stamm LA, Bush D;  
 XX

DR WPI: 2002-381255/41.  
DR N-PSDB: ABN91568.  
XX Novel isolated nucleic acid encoding a Staphylococcus epidermidis  
PT polypeptide, useful for diagnosing and treating bacterial infections -  
XX  
PS Disclosure; SEQ ID 3868; 267pp; English.  
XX  
XX ABN90538 to ABN93374 represent Staphylococcus epidermidis open reading  
CC frame (ORF) nucleic acid sequences which encode the amino acid sequences  
CC given in ABP5124 to ABP37960. The S. epidermidis sequences have  
CC antibacterial activity and can be used in gene therapy. The sequences  
CC can also be used in the diagnosis and treatment of bacterial infections,  
CC particularly S. epidermidis infections. The sequences can be used to  
CC screen for compounds able to interfere with the S. epidermidis life  
CC cycle or inhibit S. epidermidis infection.  
CC N.B. The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from the  
CC USPTO web site.  
CC  
SQ Sequence 778 AA;  
Query Match 14.8%; Score 118; DB 23; Length 778;  
Best Local Similarity 27.2%; Pred. No. 0.0067;  
Matches 47; Conservative 23; Mismatches 57; Indels 46; Gaps 9;  
QY 1 DTGAVSEIKPRVYTTIQNGKMSSTIVSEDFILPVK-----GHEKGYQFGW--- 51  
DB 618 DSVNAQSLKP-----ITINGKQIKQOQSVKSGTKVLPHSKVLMATDGEILMP-DMTGWTKB 672  
QY 52 EISGFE-----GKKDAGYVYN--LSKDTFKPVPKKEKEKEENKPTPDVS---KK 98  
DB 673 DYLAFFEDLTJKIKVSTKNGSPYTNQISIKQIIR-----NKKIEVLSAEDT 719  
QY 99 KDNPOVNHSQLNESHRKEDLOREESOKSDSTKDVATVLDKNISKSTNN 151  
DB 720 DDDDEKTDSDSSDNKSKKQKADHDHNTSSSTKN-----DKSNADSNDSDD 766  
RESULT 8  
AAG47777  
ID AAG47777 standard; Protein; 484 AA.  
XX  
AC AAG47777;  
XX  
DT 18-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 60255.  
XX  
KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX  
OS Arabidopsis thaliana.  
XX  
PN EP1033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-0301439.  
XX  
XX 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 23-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.

PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0137222.  
PR 01-JUN-1999; 99US-0137528.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137528.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139452.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0143977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.

PR	14-OCT-1999;	99US-0159637.
PR	14-OCT-1999;	99US-0159638.
PR	18-OCT-1999;	99US-0159584.
PR	21-OCT-1999;	99US-0160741.
PR	21-OCT-1999;	99US-0160767.
PR	21-OCT-1999;	99US-0160768.
PR	21-OCT-1999;	99US-0160770.
PR	21-OCT-1999;	99US-0160814.
PR	21-OCT-1999;	99US-0160815.
PR	22-OCT-1999;	99US-0160980.
PR	22-OCT-1999;	99US-0160981.
PR	22-OCT-1999;	99US-0160989.
PR	25-OCT-1999;	99US-0161404.
PR	25-OCT-1999;	99US-0161405.
PR	25-OCT-1999;	99US-0161406.
PR	26-OCT-1999;	99US-0161359.
PR	26-OCT-1999;	99US-0161360.
PR	26-OCT-1999;	99US-0161361.
PR	28-OCT-1999;	99US-0161920.
PR	28-OCT-1999;	99US-0161992.
PR	28-OCT-1999;	99US-0161993.
PR	29-OCT-1999;	99US-0162142.

Query Match	13.5%;	Score 107.5;	DB 21;	Length 484;
Best Local Similarity	20.1%;	Pred. No. 0.039;		
Matches	36;	Conservative	32;	Mismatches 66;
				Indels 45;
				Gaps 5;

```

QY 9 KPHVYVTLQNGKMSSTVSEEDFILPVYKGLKNGQPGWELSGE-----GK 60
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 96 RKNRTVDVQNNNGSESK-----YQDAPRRRYDE-EATYSQSAORIDHPNQ 143
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 61 DAGVINLSKDTFLKPPVKKLEKKEBENKPTPVSKKQK----- 101
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 144 NVGITEKAFENSPLEETSHRDNNKRLNNQKNFLPAALSSSENAVRVSFGADHKQAEVWGK 203
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 102 PQVHNSQLNE-----SHRKEDLQREHSQKSDSTQVATATVLDKNNLSKSTNNPKR 154
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 204 PMERNDDVQRTGSAKSHRKENVTYKSEKPRQDEGVKTEADQKQNNKEKKEKTESINX 262
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

PT	Proteins encoded by chromosome 2 of the human malarial parasite,
XX	
DR	WPI; 2000-365347/31.
XX	
PI	Hoffman S., Carucci D., Gardner M., Venter JC;
XX	
PA	(VENT/) VENTER J C.
PA	(GARD/) GARDNER M.
PA	(CARU/) CARUCCI D.
PA	(HOFF/) HOFFMAN S.
XX	
PR	05-NOV-1998; 98US-0107131.
XX	
PX	
PP	05-NOV-1999; 99WC-US26796.
XX	
PD	11-MAY-2000.
XX	
PN	WO200025728-A2.
XX	
OS	Plasmodium falciparum.
XX	
KW	antimalarial; malaria; protozoacide; infection; insecticide.
KW	Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine,
XX	
DE	Plasmodium falciparum chromosome 2 related protein SEQ ID NO:135.
XX	
DT	07-NOV-2000 (first entry)
XX	
AC	AAB18278;
XX	
ID	AAB18278 standard; Protein; 665 AA.
XX	
AB	RESULT 9
XX	

```

PT Plasmodium falciparum, useful as antimalarial vaccines and in the
PT diagnosis of P.falciparum infection -
XX
XX Disclosure; Page 321-322; 577bp; English.
XX
XX The present invention describes proteins and their fragments (I) encoded
CC by chromosome 2 of the human malarial parasite, Plasmodium falciparum.
CC Also described are: (1) nucleotide sequences (II) encoding (I); and (2)
CC vaccines against P. falciparum infection comprising (I) or (II).
CC (I) and (II) are useful for the development of vaccines against
CC P. falciparum infection. (I) and polyclonal antisera or a monoclonal
CC antibody raised to immunogens comprising the sequences of (I), are
CC useful in the detection of infection with P. falciparum. Furthermore,
CC (I) (especially when they are ritins or secreted or membrane proteins)
CC can aid the identification of drugs to treat or prevent P. falciparum
CC infection, or they can be used to identify drug resistance in
CC P. falciparum. Sequencing of the Plasmodium chromosome 2 and the
CC subsequent identification of proteins encoded by it will help to expand
CC our understanding of parasite biology, a process hampered by the
CC complexity of the parasitic lifecycle, and provide new targets for
CC vaccine and drug development. Parasite resistance to drugs and mosquito
CC resistance to insecticides have led to a resurgence of malaria in many
CC parts of the world, and there is a pressing need for vaccines and new
CC drugs. AAAT70078 to AAAT9287 and AAB19144 to AAB18352 represent nucleotide
CC and protein sequences given in the present invention, but which are not
CC specifically mentioned within the specification.
XX
XX Sequence 665 AA;
SQ
Query Match 13.1%; Score 105; DB 21; Length 665;
Best Local Similarity 24.4%; Pred. No. 0.11;
Matches 42; Conservative 34; Mismatches 40; Indels 56; Gaps 10
QY 1 DTGVESELKPHRT-VTIQNGKMSSTVSEEDFILPVYKGLKGYQFGWEI--SGFE 57
Db 128 EKMTINKEDIKRQNELNLQSGK-----NEGDI-----NRNGKQKQ---DISNSAE 170
QY 58 GKDAQYVINTLSKSTFFIKPVFKKIEKKE-----EENKPTPD---VSKKDNFQ 103
Db 171 NKKD-----KRGVTELEKKEKKEKESIDPHKVEENKKSDDHKVEENKKSDDHK 218
QY 104 VNHSGANSHRKEDLQR-BEHSQKSDSTQDVATVLDKKNISKSSTNNPK 154
Db 219 VEENKKSDDHKIEVKVKEHEDEDEB-----DKKEKSENNKKNDENK 261
RESULT 10
AAB61977
ID ABB61977 standard; Protein; 564 AA.
AC ABB61977;
DT 26-MAR-2002 (first entry)
XX Drosophila melanogaster polypeptide SEQ ID NO 12723.
XX Drosophila; developmental biology; cell signalling; insecticide;
XX pharmaceutical.
XX Drosophila melanogaster.
XX WO200171042-A2.
XX 27-SEP-2001.
XX 23-MAR-2001; 2001WO-US09231.
XX 23-MAR-2000; 2000US-191637P.
XX 11-JUL-2000; 2000US-0614150.
XX (PENE ) PE CORP NY.
XX
XX Venter JC, Adams M, Li PWD, Myers EW;

```

```

XX  WPI: 2001-6556860/75.
DR  N-PSDB; ABL06080.
XX
XX  New isolated nucleic acid detection reagent for detecting 1000 or more
PT  genes from Drosophila and for elucidating cell signalling and cell-cell
PT  interactions -
PS  Disclosure; SEQ ID NO 12723; 21PP + Sequence Listing; English.
XX
XX  The invention relates to an isolated nucleic acid detection reagent
CC  capable of detecting 1000 or more genes from Drosophila. The invention is
CC  useful in developmental biology and in elucidating cell signalling and
CC  cell-cell interactions in higher eukaryotes for the development of
CC  insecticides, therapeutics and pharmaceutical drugs. The invention
CC  discloses genomic DNA sequences (AB16176-AB160511), expressed DNA
CC  sequences (AB101840-AB16175) and the encoded proteins
CC  (ABB57737-ABB72072).
CC  The sequence data for this patent did not form part of the printed
CC  specification, but was obtained in electronic format directly from WIPO
CC  at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ  Sequence 564 AA;
XX
Query Match 12.7%; Score 101.5; DB 22; Length 564;
Best Local Similarity 24.5%; Pred. No. 0.19;
Matches 34; Conservative 29; Mismatches 51; Indels 25; Gaps 5;
QY 31 EDPLPYKGELEKGYPODCW-----EISGFEKKDAGYVI-----NLKDTFFIK 75
DB 78 EDLTPPLSESRFSK--VFDCWVDENHRDEHGHVDQPSGEGALDHDENHDHDEDEDER 135
QY 76 PVFKKITEKKKEKKKPT-----PDVKKKKDPPVNSQNSHRKXDLQREKHSQKSDS 129
DB 136 PLTEELEEEBEEBEEPEDEDEPADEEYEEDEBENNA--GENITAEADAEEBEEEDND 193
QY 130 TKQVATVLDKNNISKST 148
DB 194 EGTVEATVETATTEATT 212
XX
RESULT 11
ID AAE20967
XX  AAE20967 standard; Protein; 2060 AA.
XX
XX  AAE20967;
XX
DT 01-JUL-2002 (first entry)
XX
DB  Staphylococcus lugdunensis von Willebrand factor binding protein.
XX
XX  Von Willebrand factor binding protein; vWb; immunogen; antibacterial;
XX  vaccine; infection.
XX
XX  Staphylococcus lugdunensis.
XX
XX  Key Location/Qualifiers
XX  Peptide 1..47
XX  Protein /label= Signal_peptide
XX 48..2060
XX /note= "Mature von Willebrand factor binding protein"
XX
XX WO200228892-A1.
XX
XX 11-APR-2002.
XX
XX 06-APR-2001; 2001WO-SB00766.
XX
XX 04-OCT-2000; 2000SB-0003573.
XX
XX (BIOS-) BIOSTAPRO AB.
XX
XX Gus B, Fryberg L, Jacobsson K, Ahlen J, Nilsson M;

```

XX MPI: 2002-304928/34.  
DR N-PSDB; AAD33371.  
XX  
XX New von Willebrand factor binding protein from Staphylococci, useful  
PT for determining and treating staphylococcal infection -  
PS Claim 3; Page 35-41; 53pp; English.  
XX  
XX The present invention relates to von Willebrand factor binding protein or  
CC polypeptide (vwb) from Staphylococci. The vwb and immunogens of vwb are  
CC useful in vaccines to combat infections caused by Staphylococci. The  
CC invention is also useful for detection of staphylococcal infection and  
CC purifying von Willebrand factor from a complex solution. The present  
CC sequence is Staphylococcus lugdunensis vwb1 protein.  
XX  
SQ Sequence 2060 AA:  
Query Match 12.1%; Score 96.5; DB 23; Length 2060;  
Best Local Similarity 25.7%; Pred. No. 3.8;  
Matches 39; Conservative 25; Mismatches 49; Indels 39; Gaps 9;  
QY 4 EVSELPKRYVTYITQ-NGKMSSTIVSEEDFLIPYKGLKGYQDPNG--WEISGFEGKK 60  
DB 1914 DIDLK---ITIVDTNGRE-----IVPSRKGQLPP-EQFTGDPQYTGK--- 1955  
QY 61 DAGYVNLKSDFTIKYVKKIIE-----KGBENKPTDVSKKONPOVNSQLNESH 113  
DB 1956 -----IEKDGITTYIKYKVENAVPAKQLKTKGN--TQSESOFTATPVKQQLVXYAN 2006  
QY 114 RKE--DLQREHSOKSDSTQDVATVLDKNNI 143  
DB 2007 VNEGRSIEKSEHTDMVSELPTGETANKGL 2038  
DB  
RESULT 12  
AAB18272  
ID AAB18272 standard; Protein: 2500 AA.  
AC AAB18272;  
XX  
XX 07-NOV-2000 (first entry)  
DT  
XX Plasmodium falciparum chromosome 2 related protein SEQ ID NO:129.  
DE  
XX Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;  
KM antimalarial; malaria; protozoacide; infection; insecticide.  
XX Plasmodium falciparum.  
OS  
XX  
PN MO200025728-A2.  
XX  
XX 11-MAY-2000.  
PD  
XX 05-NOV-1999; 99WO-US26796.  
PF  
XX 05-NOV-1998; 98US-0107131.  
PR  
XX (HOFF/) HOFFMAN S.  
PA (CARD/) CARUCCI D.  
PA (GARD/) GARDNER M.  
PA (VENT/) VENTER J C.  
XX  
XX Hoffman S, Carucci D, Gardner M, Venter JC;  
PI MPI; 2000-365347/31.  
DR  
XX  
XX Proteins encoded by chromosome 2 of the human malarial parasite,  
PT Plasmodium falciparum, useful as antimalarial vaccines and in the  
PT diagnosis of P. falciparum infection -  
XX  
XX Disclosure, Page 302-309; 577pp; English.  
XX

CC The present invention describes proteins and their fragments (I) encoded  
CC by chromosome 2 of the human malarial parasite, Plasmodium falciparum.  
CC Also described are: (I) nucleotide sequences (II) encoding (I); and (2)  
CC vaccines against P. falciparum infection comprising (I) or (II).  
CC (1) and (II) are useful for the development of vaccines against  
CC P. falciparum infection. (I) and polyclonal antisera or a monoclonal  
CC antibody raised to immunogens comprising the sequences of (I), are  
CC useful in the detection of infection with P. falciparum. Furthermore,  
CC (1) (especially when they are refined or secreted or membrane proteins)  
CC can aid the identification of drugs to treat or prevent P. falciparum  
CC infection, or they can be used to identify drug resistance in  
CC P. falciparum. Sequencing of the Plasmodium chromosome 2 and the  
CC subsequent identification of proteins encoded by it will help to expand  
CC our understanding of parasite biology, a process hampered by the  
CC complexity of the parasite lifecycle, and provide new targets for  
CC vaccine and drug development. Parasite resistance to drugs and mosquito  
CC resistance to insecticides have led to a resurgence of malaria in many  
CC parts of the world, and there is a pressing need for vaccines and new  
CC drugs. AAA70078 to AAA70287 and AAB18144 to AAB18352 represent nucleotide  
CC and protein sequences given in the present invention, but which are not  
CC specifically mentioned within the specification.  
XX  
SQ Sequence 2500 AA:  
Query Match 12.0%; Score 96; DB 21; Length 2500;  
Best Local Similarity 26.1%; Pred. No. 5.5;  
Matches 41; Conservative 30; Mismatches 54; Indels 32; Gaps 7;  
QY 9 KHRVYVTITONGKMSSTIVSEEDFLIPYKGLKGYQDPNGWEISGFEGKKAGYVINTL 68  
DB 2173 KPYKIT---ENNK-----NEGNBILKTKYSINEKKNYDKQENCLDKDTQCVNVT 2223  
QY 69 -----SKDTFIKPYVKIIEKKKEENKPTFDVSKKONPOVNSQLNE---SHRKED 117  
DB 2224 KKKNNLNNKSPSPSNITVKLEBEKSDDKED---DKKNDTRKNNLNNKSPSPSNITVK 2280  
QY 118 LQREHSOKSDSTQDVATVLDKNNISKSTNNPK 154  
DB 2281 LREBE---KSDDKD-----DKKNDTRKNNLNNK 2308  
DB  
RESULT 13  
AAG37134  
ID AAG37134 standard; Protein: 408 AA.  
AC AAG37134;  
XX  
XX 18-OCT-2000 (first entry)  
DT  
XX Arabidopsis thaliana protein fragment SEQ ID NO: 45610.  
DE  
XX Protein identification; signal transduction pathway; metabolic pathway;  
KM hybridisation assay; genetic mapping; gene expression control; promoter;  
KM termination sequence.  
XX Arabidopsis thaliana.  
OS  
XX  
PN EP1033405-A2.  
XX  
XX 06-SEP-2000.  
PD  
XX 25-FEB-2000; 2000EP-0301439.  
PF  
XX 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126284.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR

PR	19-APR-1999;	99US-0130077.
PR	21-APR-1999;	99US-0130444.
PR	23-APR-1999;	99US-0130510.
PR	28-APR-1999;	99US-0130891.
PR	28-APR-1999;	99US-0131449.
PR	30-APR-1999;	99US-0132048.
PR	30-APR-1999;	99US-0132407.
PR	04-MAY-1999;	99US-0132484.
PR	05-MAY-1999;	99US-0132485.
PR	06-MAY-1999;	99US-0132486.
PR	06-MAY-1999;	99US-0132487.
PR	07-MAY-1999;	99US-0132863.
PR	11-MAY-1999;	99US-0134256.
PR	14-MAY-1999;	99US-0134218.
PR	14-MAY-1999;	99US-0134219.
PR	14-MAY-1999;	99US-0134221.
PR	14-MAY-1999;	99US-0134370.
PR	18-MAY-1999;	99US-0134768.
PR	19-MAY-1999;	99US-0134941.
PR	20-MAY-1999;	99US-0135124.
PR	21-MAY-1999;	99US-0135353.
PR	24-MAY-1999;	99US-0135623.
PR	25-MAY-1999;	99US-0136021.
PR	27-MAY-1999;	99US-0136392.
PR	28-MAY-1999;	99US-0136788.
PR	01-JUN-1999;	99US-0137222.
PR	03-JUN-1999;	99US-0137528.
PR	04-JUN-1999;	99US-0137502.
PR	07-JUN-1999;	99US-0137724.
PR	08-JUN-1999;	99US-0138094.
PR	10-JUN-1999;	99US-0138540.
PR	10-JUN-1999;	99US-0138847.
PR	14-JUN-1999;	99US-0139119.
PR	16-JUN-1999;	99US-0139452.
PR	16-JUN-1999;	99US-0139453.
PR	17-JUN-1999;	99US-0139492.
PR	18-JUN-1999;	99US-0139454.
PR	18-JUN-1999;	99US-0139455.
PR	18-JUN-1999;	99US-0139456.
PR	18-JUN-1999;	99US-0139461.
PR	18-JUN-1999;	99US-0139462.
PR	18-JUN-1999;	99US-0139463.
PR	18-JUN-1999;	99US-0139465.
PR	18-JUN-1999;	99US-0139750.
PR	18-JUN-1999;	99US-0139763.
PR	21-JUN-1999;	99US-0139817.
PR	22-JUN-1999;	99US-0139899.
PR	23-JUN-1999;	99US-0140353.
PR	23-JUN-1999;	99US-0140354.
PR	24-JUN-1999;	99US-0140695.
PR	28-JUN-1999;	99US-0140821.
PR	29-JUN-1999;	99US-0140989.
PR	30-JUN-1999;	99US-0141287.
PR	01-JUL-1999;	99US-0141842.
PR	01-JUL-1999;	99US-0142154.
PR	02-JUL-1999;	99US-0142055.
PR	06-JUL-1999;	99US-0142390.
PR	08-JUL-1999;	99US-0142803.
PR	09-JUL-1999;	99US-0142920.
PR	12-JUL-1999;	99US-0142977.
PR	13-JUL-1999;	99US-0143545.
PR	14-JUL-1999;	99US-0143624.
PR	15-JUL-1999;	99US-0144005.
PR	16-JUL-1999;	99US-0144085.
PR	16-JUL-1999;	99US-0144086.
PR	19-JUL-1999;	99US-0144325.
PR	19-JUL-1999;	99US-0144331.
PR	19-JUL-1999;	99US-0144332.
PR	19-JUL-1999;	99US-0144333.
PR	19-JUL-1999;	99US-0144334.
PR	19-JUL-1999;	99US-0144335.
PR	19-JUL-1999;	99US-0144336.
PR	19-JUL-1999;	99US-0144337.
PR	19-JUL-1999;	99US-0144338.
PR	19-JUL-1999;	99US-0144339.
PR	19-JUL-1999;	99US-0144340.
PR	19-JUL-1999;	99US-0144341.
PR	19-JUL-1999;	99US-0144342.
PR	19-JUL-1999;	99US-0144343.
PR	19-JUL-1999;	99US-0144344.
PR	19-JUL-1999;	99US-0144345.
PR	19-JUL-1999;	99US-0144346.
PR	19-JUL-1999;	99US-0144347.
PR	19-JUL-1999;	99US-0144348.
PR	19-JUL-1999;	99US-0144349.
PR	19-JUL-1999;	99US-0144350.
PR	19-JUL-1999;	99US-0144351.
PR	19-JUL-1999;	99US-0144352.
PR	19-JUL-1999;	99US-0144353.
PR	19-JUL-1999;	99US-0144354.
PR	19-JUL-1999;	99US-0144355.
PR	19-JUL-1999;	99US-0144356.
PR	19-JUL-1999;	99US-0144357.
PR	19-JUL-1999;	99US-0144358.
PR	19-JUL-1999;	99US-0144359.
PR	19-JUL-1999;	99US-0144360.
PR	19-JUL-1999;	99US-0144361.
PR	19-JUL-1999;	99US-0144362.
PR	19-JUL-1999;	99US-0144363.
PR	19-JUL-1999;	99US-0144364.
PR	19-JUL-1999;	99US-0144365.
PR	19-JUL-1999;	99US-0144366.
PR		

PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 26-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 11.7%; Score 93.5; DB 21; Length 408;  
Best Local Similarity 25.0%; Pred. No. 0.79; Indels 31; Gaps 8;  
Matches 39; Conservative 28; Mismatches 58;

QY 4 EVSELPK-HRVVTYIQNGKMSSTIVSEEDFILPVYVGELEKQVDFGMEISGFEKKDA 62  
DB 201 EKDPLEPKFVSAFLVYANERPAALREKNSVYBVK-----ITGEEMKQLSDKKCA 252  
QY 63 GY--VINLSKPFIKPVFKIIEKKKEENKPTDVSKKKQNPQVNHSQLNESHKEDLQ 120  
DB 253 PYKVAKKKQKETLQ-AMEYKRTKEE-----ALSGKE-----EEELKTKHQEALQM 301  
QY 121 EHSQKSDSTKQVATVLDNNISSKSTNN--PNK 154  
DB 302 LKKEKTDN-----LIKKEKATKKKKNNVDPNK 330

RESULT 14  
AAG37133  
ID AAG37133 standard; Protein; 456 AA.

AC AAG37133;

DT 18-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 45609.

KM Protein identification; signal transduction pathway; metabolic pathway;  
hybridisation assay; genetic mapping; gene expression control; promoter;  
termination sequence.

XX Arabidopsis thaliana.

OS Arabidopsis thaliana.

PN EP1033405-A2.

PD 06-SEP-2000.

PF 25-FEB-2000; 2000EP-0301439.

PR 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123180.

PR 09-MAR-1999; 99US-0123548.

PR 25-MAR-1999; 99US-0126264.

PR 29-MAR-1999; 99US-0126785.

PR 01-APR-1999; 99US-0127462.

PR 06-APR-1999; 99US-0128234.

PR 08-APR-1999; 99US-0128714.

PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130049.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 11-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134378.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138647.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.



```

PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151080.
PR 27-AUG-1999; 99US-0151086.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157563.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159283.
PR 13-OCT-1999; 99US-0159284.
PR 13-OCT-1999; 99US-0159295.

```

```

PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

```

Query Match 11.7%; Score 93.5; DB 21; Length 456;  
 Best Local Similarity 25.0%; Pred. No. 0.92; Mismatches 39; Conservative 28; Indels 31; Gaps 8;

```

QY 4 EVSELRP-HRVTVIIONGKEMSTVSEDFILPVYKGLBKGYOPDGMETISGFGKXDA 62
DB 249 EKDLKRPHPYSALVIANERARALRENKSVVERAK-----ITGEFKNLSDKKKA 300
QY 63 GY--VINISKDTFKPVFKKLEKKKEENKPTFDVSKKXKNPQVNHSQLMSHREKDIQR 120
DB 301 PEKVAKKKEKTYIQ-AMEBYKRTKEE-----ALSKKE-----EESILKLKQKQALQM 349
QY 121 EESHQKSDSTKQVATVLDKNNISKSTTN--PNK 154
DB 350 LKKEKXTDN-----LKKEKATVKKKKNENVDPNK 378

```

RESULT 15  
 AAG37132  
 ID AAG37132 standard; Protein: 476 AA.

AC AAG37132;

DT 18-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SRQ ID NO: 45608.

KW Protein identification; signal transduction pathway; metabolic pathway;  
 hybridisation assay; genetic mapping; gene expression control; promoter;  
 termination sequence.

OS Arabidopsis thaliana.

PN EP1033405-A2.

PD 06-SEP-2000.

PF 25-FEB-2000; 2000EP-0301439.

```

XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.

```

PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132456.  
PR 11-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137702.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 18-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.

PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 22-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 23-JUL-1999; 99US-0145276.  
PR 26-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147202.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149175.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151338.  
PR 01-SEP-1999; 99US-0151330.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.

```

PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

```

Query Match 11.7%; Score 93.5; DB 21; Length 476;  
 Best Local Similarity 25.0%; Pred. No. 0.98;  
 Matches 39; Conservative 28; Mismatches 58; Indels 31; Gaps 8;

```

QY 4 EVSELRP-HRYVTVTQNGKMSSTVSEEDFLPYVKELKGYQFDGWEISGFEGKQDA 62
DB 269 EKDPKPEHPVSAFLVYVNERRAALREKNSVYVAK-----ITGEWKNLSDKKA 320
QY 63 GY--VINLSKOTFIKPVFKIEKKEENKPTFDVSKKQDPQVNHSQLNESHREKEDLQR 120
DB 321 PYEKVAKKQKTYILO-AMEBKRTGER-----ALSQKKE-----EEELIKHKHQBALQM 369
QY 121 EHSQKSDSTYDVTATVLDKNNISKSSTNN--PNK 154
DB 370 LKKEKKTND-----LIKKEKATKCKGNENVDPRK 398

```

RESULT 16  
 ABG16636  
 ID ABG16636 standard; Protein; 2519 AA.

```

XX AC ABG16636;
XX DT 18-FEB-2002 (first entry)
XX DE Novel human diagnostic protein #16627.
XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX OS Homo sapiens.
XX PN WO200175067-A2.
XX PD 11-OCT-2001.
XX PF 30-MAR-2001; 2001WO-US08631.
XX PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX (HYSE-) HYSEQ INC.
XX PA Dmanac RT, Liu C, Tang YT;
XX PI
XX DR WPI; 2001-639362/73.

```

```

DR N-PSDB; AAS80823.
XX
XX PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX
XX PS Claim 20; SEQ ID No 46995; 103pp; English.
XX
XX CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAG30010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX SQ Sequence 2519 AA;

```

Query Match 11.6%; Score 92.5; DB 22; Length 2519;  
 Best Local Similarity 31.0%; Pred. No. 13;  
 Matches 22; Conservative 19; Mismatches 17; Indels 13; Gaps 3;

```

QY 68 LSKOTFIKPVFKIEKKEENKPTFDVSKKQDPQVNHSQLNESHREKEDLQRE-----E 122
DB 689 VKETKTKVP-----EDKKBEKPKPEKVAKKEDKTP1--KKEKPKKEVKEKVEIK 740
QY 123 HSQKSDSTKDV 133
DB 741 KBEKKEPKKEV 751

```

RESULT 17  
 AAB42897  
 ID AAB42897 standard; Protein; 209 AA.

```

XX AC AAB42897;
XX DT 08-FEB-2001 (first entry)
XX DE Human ORFX ORF2661 polypeptide sequence SEQ ID NO:5322.
XX KW Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
KW vulnery; antiporiatic; antiparkinsonian; nootropic; neuroprotective;
KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
KW immunostimulant; chromolytic; coagulant; vasotropic; antidiabetic;
KW hypotensive; dermatological; immunosuppressive; antileukematory;
KW antiviral; antibacterial; antifungal; antineumatic; antihypoid;
KW antinaemic; gene therapy; cancer; proliferative disorder; hypertension;
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW cholesterol ester storage; systemic lupus erythematosus; infection;
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW bone damage; cartilage damage; antiinflammatory disease; coagulation;
KW thrombosis; contraceptive.
XX
XX OS Homo sapiens.

```

PN WO20058473-A2.  
XX 05-OCT-2000.  
XX 31-MAR-2000; 2000WO-US086221.  
XX 31-MAR-1999; 99US-0127607.  
PR 02-APR-1999; 99US-0127636.  
PR 05-APR-1999; 99US-0127728.  
PR 30-MAR-2000; 2000US-0540763.  
XX  
PA (CURA-) CURAGEN CORP.  
PI Shinketsu RA, Leach M;  
XX MPI; 2000-602362/57.  
DR N-PSDB; AAC77106.  
XX  
PT Novel nucleic acids and peptides derived from open reading frame X,  
PT useful for treating e.g. cancers, proliferative disorders,  
PT neurodegenerative disorders and cardiovascular disease -  
XX  
PS Claim 11; Page 4497-4498; 5507pp; English.  
XX  
XX AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,  
CC which represent the human ORFX open reading frames 1 to 3161. The ORFX  
CC sequences have activities such as: cytoskeletal; hepatotropic; vulnery;  
CC antiproliferative; antiparkinsonian; nootropic; neuroprotective;  
CC osteoparatic; anticonvulsant; antiarthritic; immunosuppressant;  
CC immunostimulant; cardiac; thrombolytic; coagulant; vasotrophic;  
CC antidiabetic; hypotensive; dermatological; immunosuppressive;  
CC antiinflammatory; antibacterial; antiviral; antifungal; antineumatic;  
CC antihydroid; and antianemic. The sequences can be used for determining  
CC the presence of or predisposition to, or preventing or treating  
CC pathological conditions associated with an ORFX-associated disorder. The  
CC nucleic acids can be used to express ORFX proteins in gene therapy  
CC vectors. The proteins and nucleic acids may be used to treat cancers,  
CC proliferative disorders, neurodegenerative disorders, osteoarthritis,  
CC graft vs host disease, cardiovascular disease, diabetes mellitus,  
CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus  
CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,  
CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,  
CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,  
CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance  
CC coagulation; to inhibit thrombosis; and as a contraceptive.  
XX  
SQ Sequence 209 AA;  
Query Match 11.3%; Score 90.5; DB 21; Length 209;  
Best Local Similarity 28.1%; Pred. No. 0.62;  
Matches 47; Conservative 17; Mismatches 56; Indels 47; Gaps 9;  
QY 3 GEVSEIKPHRTVTITONGKMSSTIVS--EDFILPVYKLEKGYQF-----DGNB--- 52  
DB 18 GFLQQLSSGQL-----HGKSDSPNVYTEKKETALIERLTLEERKLTFFQQRSDLBERLY 72  
QY 53 -----ISFPEKKKAG-----YVINTSKOTFIKVFKEIKKEEENKPTDVSKKD 100  
DB 73 VEAKQONGKQGTGDKKKGGRSGSHRANKKSKETFLGVS-----KEFTDAKKNST 120  
QY 101 NQVNH-----SQMESHREKEDLOREHOKSD--STKQVATVADK 140  
DB 121 KEFVRHKEKIKQAKKA-VKENLKKPSDSVSKTFRRFKQTTNIPDE 166  
RESULT 18  
ID AAY44364 standard; protein; 757 AA.  
XX AAY44364;  
AC  
XX  
DT 14-MAR-2000 (first entry)  
XX

DE Human cell cycle regulation protein-5.  
XX  
XX CECRP-5; cell cycle regulation protein-5; cell proliferation;  
KW cell proliferative disease; cancer; atherosclerosis; cirrhosis;  
KW hepatitis; psoriasis; immune system disorder; allergy; asthma;  
KW acquired immune deficiency syndrome; Crohn's disease; Blast method;  
KW rheumatoid arthritis; gene therapy; Chromosomal mapping.  
XX  
XX Homo sapiens.  
XX  
FH Key  
FT Modified-site  
FT Location/Qualifiers  
FT 4  
FT /note= "Potential glycosylation site"  
FT 63  
FT /note= "Potential glycosylation site"  
FT 479  
FT /note= "Potential glycosylation site"  
FT 498  
FT /note= "Potential glycosylation site"  
FT 31  
FT /note= "Potential phosphorylation site"  
FT 43  
FT /note= "Potential phosphorylation site"  
FT 81  
FT /note= "Potential phosphorylation site"  
FT 96  
FT /note= "Potential phosphorylation site"  
FT 99  
FT /note= "Potential phosphorylation site"  
FT 130  
FT /note= "Potential phosphorylation site"  
FT 158  
FT /note= "Potential phosphorylation site"  
FT 162  
FT /note= "Potential phosphorylation site"  
FT 163  
FT /note= "Potential phosphorylation site"  
FT 184  
FT /note= "Potential phosphorylation site"  
FT 186  
FT /note= "Potential phosphorylation site"  
FT 188  
FT /note= "Potential phosphorylation site"  
FT 271  
FT /note= "Potential phosphorylation site"  
FT 277  
FT /note= "Potential phosphorylation site"  
FT 287  
FT /note= "Potential phosphorylation site"  
FT 289  
FT /note= "Potential phosphorylation site"  
FT 323  
FT /note= "Potential phosphorylation site"  
FT 343  
FT /note= "Potential phosphorylation site"  
FT 367  
FT /note= "Potential phosphorylation site"  
FT 370  
FT /note= "Potential phosphorylation site"  
FT 432  
FT /note= "Potential phosphorylation site"  
FT 488  
FT /note= "Potential phosphorylation site"  
FT 499  
FT /note= "Potential phosphorylation site"  
FT 207  
FT /note= "Potential phosphorylation site"  
FT 277  
FT /note= "Potential phosphorylation site"  
FT 298  
FT /note= "Potential phosphorylation site"  
FT 370  
FT /note= "Potential phosphorylation site"



Db 502 KEFVHHKIKAKA-A-VKENLKKFSDSVKSTFRHFKDTTNNIDE 547

RESULT 20  
ID ABP38188 standard; Protein; 442 AA.  
XX AC ABP38188;  
XX DT 24-JUL-2002 (first entry)  
XX DE Staphylococcus epidermidis ORF amino acid sequence SEQ ID NO:3033.  
XX KM Staphylococcus epidermidis, open reading frame; ORF, bacterial infection;  
XX KM antibacterial, gene therapy.  
XX OS Staphylococcus epidermidis.  
XX PN US6380370-B1.  
XX PD 30-APR-2002.  
XX PF 13-AUG-1998; 98US-0134001.  
XX PR 14-AUG-1997; 97US-055779P.  
XX PR 08-NOV-1997; 97US-064964P.  
XX PA (GENO-) GENOME THERAPEUTICS CORP.  
XX PI Doucette-Stamm LA, Bush D;  
XX DR WPI; 2002-381255/41.  
XX DR N-PSDB; ABN90733.  
XX PT Novel isolated nucleic acid encoding a Staphylococcus epidermidis  
XX PT polypeptide, useful for diagnosing and treating bacterial infections -  
XX PS Disclosure; SEQ ID 3033; 267DP; English.  
XX ABN90538 to ABN93374 represent Staphylococcus epidermidis open reading  
XX CC frame (ORF) nucleic acid sequences which encode the amino acid sequences  
XX CC given in ABP3124 to ABP37960. The S. epidermidis sequences have  
XX CC antibacterial activity and can be used in gene therapy. The sequences  
XX CC can also be used in the diagnosis and treatment of bacterial infections,  
XX CC particularly S. epidermidis infections. The sequences can be used to  
XX CC screen for compounds able to interfere with the S. epidermidis life  
XX CC cycle or inhibit S. epidermidis infection.  
XX CC N.B. The sequence data for this patent did not form part of the printed  
XX CC specification, but was obtained in electronic format directly from the  
XX CC USPTO web site.  
XX SQ Sequence 442 AA;  
Query Match 11.3%; Score 90; DB 23; Length 442;  
Best Local Similarity 21.8%; Pred. No. 2;  
Matches 49; Conservative 29; Mismatches 65; Indels 82; Gaps 11;

QY 1 DTGEVSELPKRRVT--IONGKE--WSTIVSEDFILPYKGL-ELEK----- 44  
DB 91 DT-ETNPVAPQEPVNTQIEKGDFTFEATVTVPEPKLQKLEIEKQETDLSDELOE 149  
QY 45 -----GYQPDGWEISG--FEKGKAGYVYINLSKDTFIKP 76  
DB 150 STDHSLSHLAENVVWEDGAVENGDTVNIPSG-SYDGEERDGGAGYDLIEGSGSFIQ 208  
QY 77 VKRIEKKKEENK-----PTDVS-----KKQNPQVNSQANE-- 111  
DB 209 FBEQJEGMTGDEKDVVTFPEEYHABELAGKATFTKTKVNEIKFYDVEBLANDEINELD 268  
QY 112 -----SHRKEDLOREHSOKSSTQDVATVVDKNNISKSTTN 150  
DB 269 SDAENVDEYKENIARKLSSEOKATEAENT-----EKBEAINKATEN 308

RESULT 21  
ID ABJ19106 standard; Protein; 645 AA.  
XX AC ABJ19106;  
XX DT 06-MAR-2003 (first entry)  
XX DE Pathogen specific antigen related staphylococcal protein SEQ ID No 414.  
XX KM Antibacterial; virucide; fungicide; protozoacide; cytostatic; anti-HIV;  
XX KM hyper-immune; serum-reactive; antigen; pathogen; tumour; allergen;  
XX KM auto-immunity; vaccine; staphylococcal infection; antibody; cancer;  
XX KM autoimmune disease; HIV; hepatitis.  
XX OS Staphylococcus sp.  
XX PN WO200259148-A2.  
XX PD 01-AUG-2002.  
XX PF 21-JAN-2002; 2002WO-EP00546.  
XX PR 26-JAN-2001; 2001AT-0000130.  
XX PA (CIST-) CISTEM BIOTECHNOLOGIES GMBH.  
XX PI Meinke A, Nagy E, Von Ahnen U, Klade C, Henics T, Zauner M;  
XX PI Minh DB, Vytvytska O, Etz H, Dryla A, Weichhart T, Halner M;  
XX PI Tempelmeier B;  
XX DR WPI; 2003-075410/07.  
XX PT Identifying, isolating and producing hyperimmune serum-reactive  
XX PT antigens from a pathogen, for preparing vaccine or medicament for  
XX PT treating or preventing e.g. staphylococcal infections, comprises  
XX PT providing antibody preparation -  
XX PS Example 7; Page 220; 252DP; English.  
XX CC The invention relates to a novel method for identifying, isolating and  
XX CC producing hyperimmune serum-reactive antigens from a pathogen, tumour,  
XX CC allergen, a tissue or host prone to auto-immunity, where the antigens  
XX CC are used in a vaccine, comprises providing antibody preparation from a  
XX CC plasma pool of a type of animal, or individual sera with antibodies  
XX CC against the specific pathogen, tumour, allergen, tissue or host prone to  
XX CC auto-immunity. The hyperimmune serum-reactive antigens comprising any of  
XX CC the 62 sequences of 53-2261 amino acids fully defined in the  
XX CC specification, or their hyperimmune fragments are useful for the  
XX CC manufacture of a pharmaceutical preparation, particularly a vaccine  
XX CC against staphylococcal infections or colonisation against S. aureus or S.  
XX CC epidermidis. The preparation of antibodies is useful for the manufacture  
XX CC of a medicament for treating or preventing staphylococcal infections or  
XX CC colonisation against S. aureus or S. epidermidis. The antibody  
XX CC preparations may also be used for diagnostic and imaging purposes. Other  
XX CC conditions that can be treated include cancer, autoimmune diseases or  
XX CC infections caused by viral (e.g. HIV, hepatitis A, B or C), fungal or  
XX CC protozoan pathogens. This sequence represents a staphylococcal protein  
XX CC relating to the method for identifying and producing pathogen specific  
XX CC antigens of the invention.  
XX SQ Sequence 645 AA;  
Query Match 11.3%; Score 90; DB 24; Length 645;  
Best Local Similarity 21.4%; Pred. No. 3.4;  
Matches 45; Conservative 32; Mismatches 67; Indels 66; Gaps 10;

QY 4 EVSELKPRHVT--ITONGKMSSTIVSEDFILPYKGL-E-KGYQF-----DQWEIS 54  
DB 341 KMTDLDQTKYVYVESVNNESMMIDTFVK-----PIKGTGLNKGKTYVWETTDNDYKDF 395  
QY 55 GREGKK-----DAGYVYINL-SKDTPI 74

Dd		::	::	::	::
Dc	396 MVEGQVWRTISKDAKNNTFTIIPYVEGKTLVDALIVKHVKITDIDYDQYHVRIVDKAFT	455SS			
Dy	75 KVFVKLIEKKKEENKPTFDV-----SKKKDNFOVNHSQLNESHKEDLQ----	REEHSQ 125S			
Dd	456 KANTDKSNKKGEQDNSAKKEATPATSTKSPPSPVEKESQNDSDKDNDLPSEVENDA	515S			
Dy	126 KSDSTDVLT-ATVLDRNITSKSTNNPNK 154				
Dd	516 SSESCKDTPTATKPTTGEVSSSTT--PTK 543				
Dd	RESULT 22				
Dc	ABP56879				
ID	ABP56879 standard; Protein; 654 AA.				
XX	ABP56879;				
AC					
XX					
DT	07-APR-2003 (first entry)				
XX					
DE	Staphylococcus epidermidis Krkn protein SEQ ID NO:10.				
XX					
KM	Staphylococcus aureus; Staphylococcus epidermidis; MSCRAM; antibody;				
KM	microbial surface component recognising adhesive matrix molecule;				
KM	surface protein; infection; antibacterial; antiinflammatory; vaccine;				
KM	immunosuppressive; antiarthritis; gene therapy; pneumonia; endocarditis;				
XX	septic arthritis; biomaterial related infection.				
XX					
OS	Staphylococcus epidermidis.				
XX					
PM	WO2002102829-A2.				
PD	27-DEC-2002.				
XX					
PE	17-JUN-2002; 2002WC-US19220.				
XX					
PR	15-JUN-2001; 2001US-298098P.				
XX					
PA	(INHI-) INHIBITEX INC.				
PA	(OUBE-) OUBEN ELIZABETH COLLEGE DUBLIN.				
PA	(UYPA-) UNIV PAVIA.				
Pt					
Pt	Poster TJ, Roche F, Patti JM, Hutchins JT, Hall A, Domanek P,				
Pt	Patel F, Syribeys P, Speciale P;				
PT	related infections -				
XX					
DR	N-PsDB; ABZ22303.				
XX					
PS	Claim 1; Page 29-30; 122pp; English.				
CC					
CC	The present invention describes an isolated antibody (I) that binds to a				
CC	staphylococcal surface protein selected from any of the 12 sequences of				
CC	354-2263 amino acids given in ABP56875 to ABP56886. Also described:				
CC	(1) an isolated antisera (II) comprising (I); (2) a diagnostic kit (III)				
CC	comprising (I) and means for detecting binding by (I); (3) diagnosing				
CC	(MI) an infection of Staphylococcus aureus by adding (I) to a sample				
CC	suspected of being infected with the infection, and determining if				
CC	antibodies have bound to the sample; (4) a pharmaceutical composition				
CC	(IV) for treating or preventing an infection of S. aureus comprising (I),				
CC	and a vehicle, carrier or excipient; (5) treating (M2) or preventing an				
CC	infection of S. aureus by administering (I) to a human or animal patient;				
CC	(6) producing (M3) an immunological response by administering (I) to a human				
CC	or animal an immunogenic amount of the isolated surface protein; (7) an				
CC	isolated active fragment (V) from the A domain of the Dsg4 protein; and				
CC	(8) a vaccine (VI) for treating or preventing an infection of S. aureus				
CC	comprising the surface protein in an amount effective to elicit an immune				
CC	response, and a vehicle, carrier or excipient. (I) has antiinflammatory,				
CC	antibacterial, immunosuppressive and antiarthritic activities, and can be				

	used in gene therapy. The antibody, composition and vaccine are useful
	for treating or preventing Staphylococcus aureus infection in a human or
	animal, such as pneumonia, septic arthritis, endocarditis or biomaterial
	related infections. The present sequence represents Staphylococcus
	epidemicidis KtKn protein, which is used in the exemplification of
	the present invention.
SQ	Sequence      654 AA;
Dg	Query Match                  11.3%; Score 90; DB 24; Length 654; Best Local Similarity    21.4%; Pred. No. 3.4; Matches     45; Conservative    32; Mismatches    67; Indels    66; Gaps    10
OY	4 EVSELEKPRRVTV--TIIONGKEWSSSTVSESDPLPVYKGELE-KGYOF-----DGWEIS 54 ::::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: :::
Db	350 KMTDIDQTKTYVESVEANNEHMDTFVKH-----PIKTGMANGSKTYMWHETTRDDTYKGF 404 ::::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: :::
OY	55 GFEGKK-----DAGVINL-SKOTFI 74 ::::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: :::
Db	40S MWEGGRVATISDKAKNRTTIFPVBEGKLTDALVKNHVKTIDVDGGVHMIVDEEAFI 464 ::::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: :::
OY	75 KPVPFKIKIEBKKEENKLPFDV----SKKGDPVNHSQLNESHREKDLO---REEHQ 125 ::::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: :::
Db	46S KANTDSNKKEQGDNDSAKKEATPATPSKPTSPPEKESQRDSQKODNRQLPSVEKENDA 524 ::::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: :::
OY	126 KSDSTKVDT-ATVLDKNNISSGSTTNPNPK 154 ::::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: :::
Db	52S SSBGSCKDTXPATKPTKPGEVSSSTI--PTK 552 ::::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: :::
RESULT 23	
ID	ABBS8769
XX	ABB58769 standard; Protein; 1183 AA.
AC	ABBS8769;
XX	
DT	26-MAR-2002 (first entry)
XX	
DE	Drosophila melanogaster polypeptide SEQ ID NO 3099.
XX	
KM	Drosophila: developmental biology; cell signalling; insecticide;
XX	pharmaceutical.
OS	Drosophila melanogaster.
PB	MO200171042-A2.
PD	27-SEP-2001.
PF	23-MAR-2001; 2001WO-US09231.
PR	23-MAR-2000; 2000US-191637P.
PR	11-JUL-2000; 2000US-0614150.
XX	
PA	(PEKE ) PE CORP NY.
PI	Venter JC, Adams M, Li PWD, Myers EW;
DR	WPI; 2001-656860/75.
N-P	N-PADB; ABL02872.
PT	New isolated nucleic acid detection reagent for detecting 1000 or more genes from Drosophila and for elucidating cell signalling and cell-cell interactions -
PT	
PS	Disclosure; SEQ ID NO 3099; 21dp + Sequence Listing; English.
CC	The invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from Drosophila. The invention is useful in developmental biology and in elucidating cell signalling and cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA

CC sequences (ABL01840-ABL16175) and the encoded proteins  
CC (AB857737-AB872072).  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 1183 AA;  
Query Match 11.3%; Score 90; DB 22; Length 1183;  
Best Local Similarity 24.7%; Pred. No. 7.8;  
Matches 36; Conservative 24; Mismatches 54; Indels 32; Gaps 5;  
QY 9 KPHRVVTTIQNGKMSSTIVSEEDFLIPYKGELEKGYCPDGEWISGFEKKGAGYVNL 68  
DB 310 KEVKTATADNKKD---PLTAEQSKPEKVPKKEYSK-----TGGKKGGA----- 352  
QY 69 SKDTFIPVFKIEKKKEBENKPTFDVSKKKNPOVNHQINSRKEQLQREHSQ--- 125  
DB 353 -----KPAEK--SKKEEKDSTTKSKKEADSPANNQKQIGVKKKTSPEPDASHK 404  
QY 126 ----KSDSTQDVATVLDKNNISKS 147  
DB 405 SVRAKDKSKDEASTQVKSNDNSPEA 430  
RESULT 24  
AAV03190  
ID AAV03190 standard; Protein: 402 AA.  
XX  
XX AAY03190;  
XX  
DT 16-JUN-1999 (first entry)  
XX  
DE S. aureus trigger factor protein sequence.  
XX  
XX Trigger factor; tig; diagnosis; immune response; bacterial infection;  
XX S. aureus infection; otitis media; toxic shock syndrome; conjunctivitis;  
XX wound infection; impetigo; septic arthritis; therapy.  
XX  
OS Staphylococcus aureus.  
XX  
XX EP905233-A2.  
XX  
XX 31-MAR-1999.  
XX  
XX 21-AUG-1998; 98EP-0306697.  
XX  
XX 29-DEC-1997; 97US-0899339.  
XX PR 04-SEP-1997; 97US-0057511.  
XX  
XX (SMIK ) SMITHKLINE BEECHAM CORP.  
XX PA (SMIK ) SMITHKLINE BEECHAM PLC.  
XX  
XX Burnham MKR, Fosberry A, Hodgson JE, Jaworski D;  
XX PI Lawlor EJ, Rosenberg M, Traini CM, Wang M, Ward J;  
XX PI Warren RL;  
XX  
XX WPI; 1999-192663/17.  
XX DR N-PSDB; AAX28175.  
XX  
XX  
XX New Staphylococcus aureus trigger factor (tig) polypeptide and  
XX polynucleotide, useful as diagnostic reagents and for prevention and  
XX treatment of Staphylococci aureus infections  
XX  
XX  
XX Claim 23; Page 8; 31pp; English.  
XX  
XX This sequence is the Staphylococcus aureus trigger factor (tig)  
XX polypeptide of the invention. Tig polypeptides and polynucleotides are  
XX useful for diagnosing diseases related to over or underexpression of tig  
XX protein by identifying mutations in the tig gene, or determining tig  
XX polypeptide or mRNA expression levels due to an infection of an organism  
XX with the tig gene. They can diagnose the stage and type of infection. Tig  
XX polypeptides are also useful for screening compounds which affect

CC activity of the protein by measuring the binding to tig and observing the  
CC stimulation or inhibition of the polypeptide function. These can be used  
CC in treatment to inhibit or enhance tig activity, in addition to direct  
CC administration of tig polypeptides to treat conditions associated with a  
CC lack of tig polypeptide, or direct administration of antisense sequences  
CC to prevent expression. Tig polypeptides (administered directly, in a  
CC vector and as a vaccine) and antibodies induce an immune response to  
CC immunise and prevent disease. Diseases diagnosed, prevented or treated  
CC include: bacterial infections, especially Staphylococcus aureus  
CC infections which cause otitis media, toxic shock syndrome,  
CC conjunctivitis, wound infection, impetigo and septic arthritis etc. Tig  
CC polypeptides, polynucleotides and their (ant)agonists can prevent  
CC adhesion of bacteria to matrix proteins, and are useful for use on wounds  
CC and body implants to prevent bacterial infection.  
XX  
SQ Sequence 402 AA;  
Query Match 11.2%; Score 89.5; DB 20; Length 402;  
Best Local Similarity 22.1%; Pred. No. 2;  
Matches 50; Conservative 33; Mismatches 60; Indels 83; Gaps 13;  
QY 1 DTGVSBLKPHRVVTT-IQNGKE--MSSTIVSEEDFLIPYKGELEK----- 44  
DB 53 DETDIKPVAAQPEVSVTQIEKGDPIFEATVTVPEVLDGYKGLIEKQETELSDDELOE 112  
QY 45 -----GYPDGWEISG--FEKKGAGYVNLSDQTRPK 76  
DB 113 AIDHSILGLAMVYKEDGVENGDTVINIDPSG--SYDGEEREGQAEQYDEIGSGSP1-P 170  
QY 77 VFK-----KIEKKK-----EE---NKPTFDVS---KKKNPOVNHQINSQ--- 111  
DB 171 GFEBQLEGMKVDEBKDVVTPPEEYHAEELAGKAEATFKTVNIEIKFKEVPELTDIELANEL 230  
QY 112 -----SHRKEDLQREHSQKSDSTQDVATVLDKNNISKSSTTN 150  
DB 231 DAAEATVDEYKENLRKRLAQKATDAENV-----EKSEATTKATDN 271  
RESULT 25  
AAV35091  
ID AAV35091 standard; Protein: 511 AA.  
XX  
XX AAY35091;  
XX  
XX 13-SEP-1999 (first entry)  
XX  
XX Chlamydia pneumoniae transmembrane protein sequence.  
XX  
XX Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;  
XX sinusitis; purulent otitis media; erythema nodosum; pharyngitis;  
XX vaccine; neutralising epitope.  
XX  
XX Chlamydia pneumoniae.  
XX OS  
XX  
XX WO9927105-A2.  
XX  
XX 03-JUN-1999.  
XX  
XX 20-NOV-1998; 98WO-1801890.  
XX  
XX 04-NOV-1998; 98US-0107078.  
XX PR 21-NOV-1997; 97FR-0014673.  
XX  
XX (GEST ) GENSET.  
XX PA  
XX  
XX Griffiths R;  
XX  
XX WPI; 1999-357842/30.  
XX  
XX Genome sequence of Chlamydia pneumoniae  
XX  
XX  
XX Page 975-976; Disclosure; 1912pp; English.  
XX





PT	interactions -
XX	
PS	Disclosure; SEQ ID NO 2904; 21bp + Sequence Listing; English.
XX	
CC	The invention relates to an isolated nucleic acid detection reagent
CC	capable of detecting 1000 or more genes from Drosophila. The invention is
CC	useful in developmental biology and in elucidating cell signalling and
CC	cell-cell interactions in higher eukaryotes for the development of
CC	insecticides, therapeutics and pharmaceutical drugs. The invention
CC	discloses genomic DNA sequences (AB16176-AB130511), expressed DNA
CC	sequences (AB101840-AB16175) and the encoded proteins
CC	(ABB57737-ABB72072).
CC	The sequence data for this patent did not form part of the printed
CC	specification, but was obtained in electronic format directly from WIPO
CC	at ftp.wipo.int/pub/published_pct_sequences.
XX	
SQ	Sequence 1408 AA;
Query Match	11.2%; Score 89.5; DB 22; Length 1408;
Best Local Similarity	24.7%; Pred. No. 11;
Matches 48; Conservative 23; Mismatches 62; Indels 61; Gaps 9	
Dd	
19 NGKR-----MSSTVSEDPF--ILPYKGELEKGYPDGMELISFGCK--KDAGYV----- 65	
646 NKKQLNASKETKTETBKDFEIVSKKE--EKNDGNSLNLNQPKLKPDQDYLDVKRP 704	
Dy	
66 -----INLSKDTFIKVPK---KIEKKKEENKPFTDVSKKKDNPDVNHSQLN 110	
:               :               :               :               :	
705 MESKEYSEPLAVKLGEDVSKSPSNSSDPTKISEVGEE--PRNGELPMTGBQVQPEKS 762	
Dz	
111 ESHRK-EIDLREHSHQKSDSTDVT-----ATVLDK 140	
::	
763 KTKAQIKIEFLAKVQSELKSKSVADVTPKQTYTEETHKGFSPKREATPKSEBPGTTTVDK 822	
Dq	
141 NNISKSTTNPNK 154	
:                :                :                :                :	
Dk	
823 SNDVSVIKSEPDE 836	
RESULT 28	
ABJ18979	
ID	ABJ18979 standard; Protein; 645 AA.
XX	
AC	ABJ18979;
XX	
DT	06-MAR-2003 (first entry)
XX	
DE	Pathogen specific antigen related staphylococcal protein SEQ ID NO 142.
XX	
KM	Antibacterial; virucide; fungicide; protozoacide; cytostatic; anti-HIV;
KM	hyperimmune; serum-reactive; antigen; pathogen; tumour; allergen;
KM	auto-immunity; vaccine; staphylococcal infection; antibody; cancer;
XX	autoimmune disease; HIV; hepatitis.
OS	Staphylococcus sp.
XX	
PN	WO200259148-A2.
XX	
PD	01-AUG-2002.
XX	
Pf	21-JAN-2002; 2002MO-EP00546.
XX	
PR	26-JAN-2001; 2001AT-0000130.
XX	
PA	(CIST-) CISTEM BIOTECHNOLOGIES GMBH.
XX	
PI	Weinke A, Nagy B, Von Ahnen U, Klade C, Henics T, Zauner W,
PI	Mint D, Vytvytska O, Etz H, Dryla A, Weichart T, Hafner M,
PI	Temelmaier B;
XX	
DR	WPI; 2003-075410/07.
XX	
PT	Identifying, isolating and producing hyperimmune serum-reactive

```

PT antigens from a pathogen, for preparing vaccine or medicament for
PR treating or preventing e.g. staphylococcal infections, comprises
PI providing antibody preparation -
XX
XX Claim 21; Page 168; 252pp; English.
XX
XX The invention relates to a novel method for identifying, isolating and
CC producing hyperimmune serum-reactive antigens from a pathogen, tumour,
CC allergen, a tissue or host prone to auto-immunity, where the antigens
CC are used in a vaccine, comprises providing antibody preparation from a
CC plasma pool of a type of animal, or individual sera with antibodies
CC against the specific pathogen, tumour, allergen, tissue or host prone to
CC auto-immunity. The hyperimmune serum-reactive antigens comprising any of
CC the 62 sequences of 53-261 amino acids fully defined in the
CC specification, or their hyperimmune fragments are useful for the
CC manufacture of a pharmaceutical preparation, particularly a vaccine
CC against staphylococcal infections or colonisation against S. aureus or S.
CC epidermidis. The preparation of antibodies is useful for the manufacture
CC of a medicament for treating or preventing staphylococcal infections or
CC colonisation against S. aureus or S. epidermidis. The antibody
CC preparations may also be used for diagnostic and imaging purposes. Other
CC conditions that can be treated include cancer, autoimmune diseases or
CC infections caused by viral (e.g. HIV, hepatitis A, B or C), fungal or
CC protozoan pathogens. This sequence represent a staphylococcal protein
CC relating to the method for identifying and producing pathogen specific
CC antigens of the invention.
CX
SQ Sequence 645 AA;
Query Match 11.1%; Score 89; DB 24; Length 645;
Best Local Similarity 21.4%; Pred. No. 4.2;
Matches 45; Conservative 32; Mismatches 67; Indels 66; Gaps 10;
QY EVSELEKPHRYTV--TIONGKMESSTIVEEDFILPVYKGBLE-XGYOF-----DGWEIS 54
   ::::| | | | | ::|| | | | | | | | | | | || | | | |
DB KATLODTKYVYSVENNESMOTFYKH-----PITGMNGKKGVWMTETNDYDKF 395
   55 GFEEKKD-----DAGYVINL-SKDTFI 74
      ||::| | | | | | | | | | | | | | | | | | | |
DB MVBDGRVRTTISKDAKNTRTIIFPYBEGKTLYDALIVKVHVTKTIDYDOGYNRIYDKCAFT 455
       75 KPVFKEIEEKEENKPFDV---SKKDNPVNHSQSLESRRKEDLQ---REBSHQ 125
        456 KANTDNKSKEGGQNNSAKKETPATPSKPTRSPVEKSESQKDSQKDNKLPLSVYEKENDA 515
          126 KSDSTKOYT-A TVLDKNNISSKSTTNPNPK 154
            | | | | | | | | | | | | | | | | | | | |
           516 SSBSGKGYTLATKPTKGEVESSTT-P TK 543
                _____
RESULT 29
ABBE4828
ID ABBE4828 standard; Protein; 281 AA.
XX
AC ABBE4828;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 21276.
XX
FM Drosophilae developmental biology; cell signalling; insecticide;
XX pharmacaceutical.
OS Drosophila melanogaster.
XX
PX MO200171042-AA.
PD
PF 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-USO9231.
XX
PF 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
```

```

XX (PEKE ) PE CORP NY.
XX
XX Venter JC, Adams M, Li FMD, Myers EW;
XX
XX WPI; 2001-656860/75.
XX
XX N-PSDB; ABL08931.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from Drosophila and for elucidating cell signalling and cell-cell
XX interactions -
XX
XX Disclosure; SEQ ID NO 21276; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signalling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (AB16176-AB160511), expressed DNA
XX sequences (AB101840-AB16175) and the encoded proteins
XX (AAB57737-AB572072).
XX
XX The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 281 AA;
SQ
Query Match 11.1%; Score 88.5; DB 22; Length 281;
Best Local Similarity 22.1%; Pred. No. 1.5;
Matches 31; Conservative 34; Mismatches 52; Indels 23; Gaps 5;
OY 19 NGKMSSTVSESEFILPVYKGELEKGYQFGWELSGEGKKDAGYV-----NLS 69
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 137 NFEFTMMWTTOQDLKNLY-SLDDK-----DNEGATSKELGAVTRALGRQPNESIA 188
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 70 KDFPKVFPKKEE-KKEEENKPTFEDVSKKKNPQVNHSQLNE-----SHRKEDLOREEH 123
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 189 KEFCNVILRMDHNTNBEELRDAPRVFKNENGYITTELRPAVFMALGRLEDELEEM 248
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 124 SOKSDSTKDVATVLDKNNI 143
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 249 IREYDLDDDNHINPESSNM 268
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
RESULT 30
ID AAG85008 standard; Protein; 1141 AA.
XX
XX AAG85008;
XX
XX 11-SEP-2001 (first entry)
XX
XX Shrimp white spot Bacilliform virus (WSBV) protein 99.
XX
XX Shrimp white spot Bacilliform virus; WSBV; diagnosis; viral infection;
XX antiviral agent; gene expression; antisense construct;
XX transgenic viral resistant shrimp.
XX
XX White spot syndrome virus.
XX
XX WO200138351-A2.
XX
XX 31-MAY-2001.
XX
XX 08-NOV-2000; 2000WO-US28888.
XX
XX 24-NOV-1999; 99CN-0124717.
XX
XX (PENY-) PE CORP NY.
XX
XX (THIR-) THIRD INST OCEANOGRAPHY STATE OCEANI C A.
XX (SINO-) SINOGENOMAX CO LTD.
XX

```

```

PI Xu X, Yang F, He J, Pham L, He M, Ye Y, Shen Y, Kodira C;
XX
XX WPI; 2001-355877/37.
XX
XX N-PSDB; AAH62788.
XX
XX Primary nucleotide sequence of the shrimp white spot Bacilliform virus
XX (WSBV), useful for producing viral polypeptides that can be used to
XX screen for agents that are useful for treating WSBV infection -
XX
XX Claim 1; Figure 3; 626pp; English.
XX
XX The invention provides the primary nucleotide sequence of the WSBV genome
XX (AAH62689), predicted transcript sequences (AAH62689-AAH62839) and
XX encoded proteins (AAG84910-AAG85051) and oligonucleotide sequences
XX (AAH62840-63160) suitable for use as primers or probes. The nucleic acid
XX molecules and proteins of the invention are useful for diagnosis and
XX monitoring viral infection. In screens for antiviral agents and for
XX monitoring viral gene expression or activity during a treatment regimen.
XX The nucleic acid molecules are also useful as antisense constructs to
XX control viral gene expression in infected cells and tissues and to create
XX transgenic viral resistant shrimp.
XX
XX Sequence 1141 AA;
SQ
Query Match 11.1%; Score 88.5; DB 22; Length 1141;
Best Local Similarity 23.3%; Pred. No. 11;
Matches 30; Conservative 31; Mismatches 43; Indels 25; Gaps 4;
OY 25 STVSEDFILPVYKGELEKGYQFGWELSGEGKKDAGYVIMLSOTPIKVPKKE 82
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 621 SNVEEEB-----EEQMEEBEVEEBGSDKDDGDA-----PQBEKE 662
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 83 EKKEENKPTFDSKKNPQVNHSQLNESHRKEDLOREEHSOKSDSTKDVATVLDKNN 142
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 663 EEKEEBQ-----QQPEESNGENQEBEQQQQPPKEEKNKADSDSDSSSSSS 717
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 143 ISSKSTTN 151
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 718 SSSSSSSS 726
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
RESULT 31
ID ABP73992 standard; Protein; 225 AA.
XX
XX ABP73992;
XX
XX 30-JAN-2003 (first entry)
XX
XX Candida albicans essential protein SEQ ID NO 7829.
XX
XX Fungus; Yeast; tetracyclin; promoter; GRACE strain; biosynthesis;
XX signal transduction; DNA replication; cell division; growth;
XX proliferation; Candida albicans; fungicide; antifungal.
XX
XX Candida albicans.
XX
XX WO200253728-A2.
XX
XX 11-JUL-2002.
XX
XX 26-DEC-2001; 2001WO-US49486.
XX
XX 29-DEC-2000; 2000US-259128P.
XX 20-FEB-2001; 2001US-0792024.
XX 22-AUG-2001; 2001US-314050P.
XX
XX (ELIT-) ELITRA PHARM INC.
XX
XX Roemer T, Jiang B, Boone C, Bussey H, Ohlsen KL;
XX
XX WPI; 2002-566994/60.
XX
XX N-PSDB; ABZ32542.
XX

```

XX	Constructing strains for identifying gene products as effective targets
PT	for therapeutic intervention, by inactivating in the strain one allele
PT	of a gene and placing other allele of the gene under conditional
PT	expression -
XX	
XX	Claim 44; SEQ ID NO 7829; 167pp + Sequence Listing; English.
PS	
CC	The invention relates to constructing (M1) a strain of diploid fungal
CC	cells in which both alleles of a gene are modified, comprising modifying
CC	one allele by insertion or replacement by a cassette having an
CC	expressible selectable marker and modifying other allele by
CC	recombination, of a promoter replacement fragment with a heterologous
CC	promoter, so that expression of the second allele is regulated by the
CC	promoter. (M1) is useful for constructing a strain of diploid fungal
CC	cells in which both alleles of a gene are modified. The diploid fungal
CC	cells having both alleles modified are useful for identifying a gene that
CC	is essential to the survival or growth of a fungus, a gene that
CC	contributes to the virulence and/or pathogenicity of a fungus, a gene
CC	that contributes to the resistance of a diploid fungus to an antifungal
CC	agent, an antifungal agent that inhibits the growth of a diploid fungus
CC	and for identifying a therapeutic agent for treatment of a mammalian
CC	disease. (M1) is useful for identifying a compound which modulates the
CC	activity of a gene product, preferably enzymatic activity, carbon
CC	compound catabolism, biosynthetic, transporter, transcriptional,
CC	translational, signal transduction, DNA replication and cell division
CC	activity. The method is useful for identifying a compound having the
CC	ability to inhibit growth or proliferation of C. albicans cells and for
CC	treating infection by C. albicans. The present sequence is that of an
CC	essential Candida albicans protein used in the method of the invention.
CC	Note: The sequence data for this patent is not represented in the printed
CC	specification but is based on sequence information supplied to Derwent by
CC	the European Patent Office.
XX	
SQ	Sequence 225 AA:
XX	
Query Match	11.0%; Score 88; DB 23; Length 225;
Best Local Similarity	28.0%; Pred. No. 1.2;
Matches	30; Conservative 22; Mismatches 43; Indels 12; Gaps 4;
QY	46 YQFQGWELSGREGKQAGVYVNLSCDTYIKPFKK----IEKKKEENKPTFDVSKK-- 99
DB	80 YDDDDDEEGFESSNGAKKEINTLSSQAIKEWQQRDLIEIEREKLNKKKEBIIIEKAKS 139
QY	100 --DNPQVHSQLNESHRKEDLQREH--SQKSDSKQVATATLIDKON 142
DB	140 TIDDFEYENSKRDNHQKEIILSEQKFTISKDDFLK--RGTLMDRYN 184
RESULT 32	
AAB94584	
ID	AAB94584 standard; Protein; 258 AA.
XX	
XX	AAB94584;
XX	
DT	26-JUN-2001 (first entry)
XX	
DE	Human protein sequence SEQ ID NO:15383.
XX	
XX	Human; primer; detection; diagnosis; antisense therapy; gene therapy.
OS	Homo sapiens.
XX	
PN	EP1074617-A2.
XX	
PD	07-FEB-2001.
XX	
PF	28-JUL-2000; 2000EP-0116126.
XX	
PR	29-JUL-1999; 99JP-0248036.
PR	27-AUG-1999; 99JP-0300253.
PR	11-JAN-2000; 2000JP-0118776.
PR	02-MAY-2000; 2000JP-0183767.

[illegible]

PD 09-AUG-2001.  
XX  
XX 05-FEB-2001; 2001WO-US04098.  
XX  
XX 03-FEB-2000; 2000US-0496914.  
PR 27-APR-2000; 2000US-0560875.  
PR 20-JUN-2000; 2000US-0598075.  
PR 19-JUL-2000; 2000US-0620325.  
PR 01-SEP-2000; 2000US-0654936.  
PR 15-SEP-2000; 2000US-0663561.  
PR 20-OCT-2000; 2000US-0693325.  
PR 30-NOV-2000; 2000US-0728422.  
XX  
XX (HYSE-) HYSEQ INC.  
XX PA  
XX Tang YT, Liu C, Dmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;  
PI Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;  
PI Xue AJ, Yang Y, Wejrtman T, Goodrich R;  
XX  
XX MPI; 2001-476283/51.  
DR N-PSDB; AAK52451.  
XX  
XX Nucleic acids encoding polypeptides with cytokine-like activities,  
PT useful in diagnosis and gene therapy -  
XX  
XX Claim 20; Page 215; 6221pp; English.  
XX  
XX The invention relates to polynucleotides (AAK51456-AAK53435) and the  
CC encoded polypeptides (AAM78323-AAM80302) that exhibit activity elating to  
CC cytokine, cell proliferation or cell differentiation or which may induce  
CC production of other cytokines in other cell populations. The  
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
CC peptide therapy. The polypeptides have various cytokine-like activities,  
CC e.g. stem cell growth factor activity, haematopoiesis regulating  
CC activity, tissue growth factor activity, immunomodulatory activity and  
CC activin/inhibin activity and may be useful in the diagnosis and/or  
CC treatment of cancer, leukemia, nervous system disorders, arthritis and  
CC inflammation.  
CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666  
CC (AAM80020) are omitted as the relevant pages from the sequence listing  
CC were missing at the time of publication.  
XX  
XX Sequence 817 AA;  
SQ

Query Match 11.0%; Score 88; DB 22; Length 817;  
Best Local Similarity 27.3%; Pred. No. 7.4;  
Matches 33; Conservative 21; Mismatches 55; Indels 12; Gaps 4;  
XX  
XX 40 GLELEKGYQPDGMEISG--FEQKQAGYVINSKDTFIPVFKIEBKEEENKPTFVSK 97  
DB 196 GQEKKQSSFKSWBSGKHQEVSKPAVSLQKQDTSKLKSTLPBQKKQESKSKPSPSQ 255  
QY 98 -KDONPVNHSOLNESHKED-----LQREHSQKSDSTKQVATATVLDKNNISSKST 148  
DB 256 WKQDTPKSKAGYVEHKKQETPKLMPVQQLKEQ-DPKKQPKSWTSMQSEQNTTTSWT 314  
QY 149 T 149  
DB 315 T 315

RESULT 34  
AAM79319  
ID AAM79319 standard; Protein; 817 AA.  
XX  
XX AAM79319;  
AC  
XX 06-NOV-2001 (first entry)  
XX  
XX Human protein SEQ ID NO 2965.  
DE  
XX  
XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;

KW tissue growth factor; immunomodulatory; cancer; leukemia;  
KW nervous system disorder; arthritis; inflammation.  
XX  
XX Homo sapiens.  
OS  
XX WO200157190-A2.  
PN  
XX  
XX 09-AUG-2001.  
PD  
XX  
XX 05-FEB-2001; 2001WO-US04098.  
XX  
XX 03-FEB-2000; 2000US-0496914.  
PR 27-APR-2000; 2000US-0560875.  
PR 20-JUN-2000; 2000US-0598075.  
PR 19-JUL-2000; 2000US-0620325.  
PR 01-SEP-2000; 2000US-0654936.  
PR 15-SEP-2000; 2000US-0663561.  
PR 20-OCT-2000; 2000US-0693325.  
PR 30-NOV-2000; 2000US-0728422.  
XX  
XX (HYSE-) HYSEQ INC.  
XX PA  
XX Tang YT, Liu C, Dmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;  
PI Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;  
PI Xue AJ, Yang Y, Wejrtman T, Goodrich R;  
XX  
XX MPI; 2001-476283/51.  
DR N-PSDB; AAK52452.  
XX  
XX Nucleic acids encoding polypeptides with cytokine-like activities,  
PT useful in diagnosis and gene therapy -  
XX  
XX Claim 20; Page 215; 6221pp; English.  
XX  
XX The invention relates to polynucleotides (AAK51456-AAK53435) and the  
CC encoded polypeptides (AAM78323-AAM80302) that exhibit activity elating to  
CC cytokine, cell proliferation or cell differentiation or which may induce  
CC production of other cytokines in other cell populations. The  
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
CC peptide therapy. The polypeptides have various cytokine-like activities,  
CC e.g. stem cell growth factor activity, haematopoiesis regulating  
CC activity, tissue growth factor activity, immunomodulatory activity and  
CC activin/inhibin activity and may be useful in the diagnosis and/or  
CC treatment of cancer, leukemia, nervous system disorders, arthritis and  
CC inflammation.  
CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666  
CC (AAM80020) are omitted as the relevant pages from the sequence listing  
CC were missing at the time of publication.  
XX  
XX Sequence 817 AA;  
SQ

Query Match 11.0%; Score 88; DB 22; Length 817;  
Best Local Similarity 27.3%; Pred. No. 7.4;  
Matches 33; Conservative 21; Mismatches 55; Indels 12; Gaps 4;  
XX  
XX 40 GLELEKGYQPDGMEISG--FEQKQAGYVINSKDTFIPVFKIEBKEEENKPTFVSK 97  
DB 196 GQEKKQSSFKSWBSGKHQEVSKPAVSLQKQDTSKLKSTLPBQKKQESKSKPSPSQ 255  
QY 98 -KDONPVNHSOLNESHKED-----LQREHSQKSDSTKQVATATVLDKNNISSKST 148  
DB 256 WKQDTPKSKAGYVEHKKQETPKLMPVQQLKEQ-DPKKQPKSWTSMQSEQNTTTSWT 314  
QY 149 T 149  
DB 315 T 315

RESULT 35  
ABP73209  
ID ABP73209 standard; Protein; 635 AA.  
XX  
XX ABP73209;  
AC

XX	30-JAN-2003	(first entry)	
DT			
DE	Candida albicans essential protein SEQ ID NO 7046.		
XX			
XX	Fungus; yeast; tetracycline promoter; GRACE strain; biosynthesis;		
KM	signal transduction; DNA replication; cell division; growth;		
KM	proliferation; Candida albicans; fungicide; antifungal.		
XX			
XX	Candida albicans.		
XX			
XX	WO200253728-A2.		
PD			
PD	11-JUL-2002.		
PE			
XX	26-DEC-2001; 2001WO-US49486.		
XX			
PR	29-DEC-2000; 2000US-259128P.		
PR	20-FEB-2001; 2001US-0792024.		
PR	22-AUG-2001; 2001US-314050P.		
XX			
PA	(ELIT-) ELITRA PHARM INC.		
P1			
XX	Roemer T, Jiang B, Boone C, Bussey H, Ohlsen KL;		
XX	MP1: 2002-566694/60.		
DR	N-PSDB; AB231759.		
XX			
PT	Constructing strains for identifying gene products as effective targets		
PT	for therapeutic intervention, by inactivating in the strain one allele		
PT	of a gene and placing other allele of the gene under conditional		
PT	expression -		
XX			
PS	Claim 44; SEQ ID NO 7046; 167pp + Sequence Listing; English.		
XX			
CC	The invention relates to constructing (M1) a strain of diploid fungal		
CC	cells in which both alleles of a gene are modified, comprising modifying		
CC	one allele by insertion or replacement by a cassette having an		
CC	expressible selectable marker and modifying other allele by		
CC	recombination, of a promoter replacement fragment with a heterologous		
CC	promoter, so that expression of the second allele is regulated by the		
CC	promoter. (M1) is useful for constructing a strain of diploid fungal		
CC	cells in which both alleles of a gene are modified. The diploid fungal		
CC	cells having both alleles modified are useful for identifying a gene that		
CC	is essential to the survival or growth of a fungus, a gene that		
CC	contributes to the virulence and/or pathogenicity of a fungus, a gene		
CC	that contributes to the resistance of a diploid fungus to an antifungal		
CC	agent, an antifungal agent that inhibits the growth of a diploid fungus		
CC	and for identifying a therapeutic agent for treatment of a mammalian		
CC	disease. (M1) is useful for identifying a compound which modulates the		
CC	activity of a gene product, preferably enzymatic activity, carbon		
CC	compound catabolism, biosynthetic, transporter, transcripional,		
CC	translational, signal transduction, DNA replication and cell division		
CC	activity. The method is useful for identifying a compound having the		
CC	ability to inhibit growth or proliferation of C. albicans cells and for		
CC	treating infection by C. albicans. The present sequence is that of an		
CC	essential Candida albicans protein used in the method of the invention.		
CC	Note: The sequence data for this patent is not represented in the printed		
CC	specification but is based on sequence information supplied to Derwent by		
CC	the European Patent Office.		
XX			
XX	Sequence 635 AA;		
Q0			
Query Match	11.0%; Score 87.5; DB 23; Length 635;		
Best Local Similarity	22.1%; Pred. No. 5.9;		
Matches	36; Conservative 38; Mismatches 68; Indels 23; Gaps 7		
QY	4 EVSLKPH--RVTVTTONGKEM-----SSTIVSEDEFLPYKGLKRGYQFDGM--- 51		
DB	70 KMKKKRPSPPWMTMPTPKGGIELNHWKSKSELBOQE---EBDGTPEYFFKXNQ 125		
QY	52 -EISGF--EGKDDAGVNLK--DFFIKVFFKIKKKEENKPTFDVSKKDDPNVYNS 107		

Db	126	LEIFPVEDVDYD..YMEIEKYESGMEKERERARERKERERKRDLEKKKKQ0000KS	184
Qy	108	QINESHRKBDLQREHSQKSDSTKQVATVLDKNNISSTTN	150
Db	185	QONPQNIKDEKNQDTRNNTGDKDSEKSDKPTVEAKKETD	227
RESULT 36			
ABBI2281			
ID	ABBI2281	standard; peptide; 2515 AA.	
AC	ABBI2281;		
XX			
DT	11-JAN-2002	(first entry)	
XX			
DE	Human secreted protein homologue, SEQ ID NO:2651.		
XX			
KW	Human; cytokine; cell proliferation; cell differentiation; growth factor;		
KW	haematopoiesis regulation; tissue growth; immunomodulator; activin;		
KW	inhibin; chemotaxis; chemokinesis; thrombolysis; oncogenesis;		
KW	proliferation; metastasis; cancer; tumour; haematopoietic disorder;		
KW	myeloid cell disorder; lymphoid cell disorder; asthma; arthritis;		
KW	chronic inflammatory condition; proliferative retinopathy;		
KW	atherosclerosis; coronary heart disease; arterial ischaemia;		
KW	bone disorder; osteoporosis; vascular growth disorder;		
KW	tissue regeneration; wound healing; infection; immune disorder;		
KW	cell culture; drug screening; gene therapy; antiinflammatory;		
KW	antiaesthetic; antiarthritic; haemostatic; antiarteriosclerotic;		
KW	cystostatic; osteopapitic; vasotrophic; cardiant; virucide; antibacterial;		
KW	antifungal; antinevray; antiulcer.		
XX			
OS	Homo sapiens.		
XX			
PM	WO200157188-A2.		
XX			
PD	09-AUG-2001.		
XX			
PF	05-FEB-2001; 2001WO-US03800.		
XX			
PR	03-FEB-2000; 2000US-0496914.		
XX			
PR	27-APR-2000; 2000US-0560875.		
XX			
PA	(HYSE-) HYSEQ INC.		
XX			
PI	Tang YT, Liu C, Dirmnac RT;		
XX			
DR	WPI; 2001-457740/49.		
XX			
DR	N-PSDB; ABA09525.		
XX			
PT	Human proteins and DNA encoding sequences useful for preventing,		
PT	treating or ameliorating a medical condition in a mammalian subject		
XX	e.g. arthritis and cancer -		
PS	Claim 20; Page 323-325; 1963jp; English.		
XX			
CC	Sequences ABBI0981-ABBI2330 represent 1350 novel human polypeptides, and		
CC	sequences ABA08225-ABA09574 represent nucleic acids encoding them. The		
CC	invention also relates to vectors and recombinant host cells comprising a		
CC	nucleotide of the invention, methods of producing the novel polypeptides,		
CC	antibodies against the polypeptides, methods of detecting the nucleotides		
CC	or polypeptides in a sample, and methods of identifying compounds which		
CC	bind to polypeptides of the invention. Although novel, many of the		
CC	polypeptides of the invention have homology to known proteins, thereby		
CC	giving an insight into their probable biological activities, and hence		
CC	potential therapeutic applications. The polypeptides of the invention may		
CC	have various activities, including cytokine, cell proliferation or cell		
CC	differentiation activities; stem cell growth factor activity;		
CC	haematopoiesis regulatory activity; tissue growth activity;		
CC	immunomodulatory activity; activin- or inhibin-related activities;		
CC	chemoattract or chemokinetic activities; haemostatic, thrombotic or		
CC	thrombolytic activities; receptor or ligand activities; or may be		
CC	involved in oncogenesis, cancer cell proliferation or metastasis.		
CC	Depending on their biological activities, polypeptides and nucleotides of		

CC the invention are useful for preventing, treating or ameliorating medical  
CC conditions, e.g., by protein or gene therapy. Such conditions include  
CC cancers, haematopoietic disorders (e.g., myeloid or lymphoid cell  
CC disorders), chronic inflammatory conditions (e.g., asthma or arthritis),  
CC proliferative retinopathy, atherosclerosis, coronary heart disease,  
CC arterial ischaemia, bone disorders (e.g., osteoporosis), and abnormal  
CC vascular growth. Polypeptides involved with tissue regeneration and  
CC repair (or nucleic acids encoding them) may be used to promote wound  
CC healing (e.g., of burns, incisions and ulcers), while those with  
CC immunomodulatory activities may be used in the treatment of viral,  
CC bacterial and fungal infections in addition to immune disorders.  
CC Polypeptides with growth factor activity may be used in cell cultures to  
CC promote cell growth. For example, such polypeptides may be used to  
CC manipulate stem cells in culture to give rise to neuroepithelial cells  
CC that can be used to augment or replace cells damaged by illness,  
CC autoimmune disease or accidental damage. The polypeptides and nucleotides  
CC may also be used in the diagnosis of the above conditions, and in drug  
CC screening techniques. The present sequence represents a novel human  
CC polypeptide of the invention.

XX Sequence 2515 AA;

Query Match 11.0%; Score 87.5; DB 22; Length 2515;  
Best Local Similarity 24.2%; Pred. No. 40;  
Matches 38; Conservative 29; Mismatches 55; Indels 35; Gaps 6;

QY 5 VSELKPRVTVYTIQNGKMSSTVSEDFILPVYKGLGKGYQDGMWISGFEKQKAGY 64

DB 2352 IAEIQREPLVNSLNVNSGFTNEIEHSYKGEISGRKQNAARISGSHVEADP-- 2409

QY 65 VINLSKDTFKPVKKEIEKKKEENKPTFDVSKKK-----DNPQVNSQLNESH 113

DB 2410 -----KEYEE--BEHMPK---RKRKQYLSSEDEPDNPVLDRI-ETA 2449

QY 114 RKEDLQREHSQKSDSTKDTATVLDKNISKSSTN 150

DB 2450 QRCPEPTEPHATKEENSRDLIE--LPKTSSETNSTS 2484

RESULT 37

AAM80268

ID AAM80268 standard; Protein: 2515 AA.

XX AAM80268;

XX 06-NOV-2001 (first entry)

XX Human protein SEQ ID NO 3914.

XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;

KM vaccine; peptide therapy; stem cell growth factor; haematopoiesis;

KM tissue growth factor; immunomodulatory; cancer; leukaemia;

XX nervous system disorder; arthritis; inflammation.

XX Homo sapiens.

XX MO200157190-A2.

XX 09-AUG-2001.

XX 05-FEB-2001; 2001WO-US04098.

XX 03-FEB-2000; 2000US-0496914.

XX 27-APR-2000; 2000US-0560875.

XX 20-JUN-2000; 2000US-0598075.

XX 19-JUL-2000; 2000US-0620325.

XX 01-SEP-2000; 2000US-0654936.

XX 15-SEP-2000; 2000US-0663561.

XX 20-OCT-2000; 2000US-0693325.

XX 30-NOV-2000; 2000US-0728422.

XX (HYSE-) HYSEQ INC.

PI Tang YT, Liu C, Dermanac RT, Aundi V, Zhou P, Xu C, Cao Y, Ma Y;  
PI Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;  
PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;  
XX WPI: 2001-476283/51.  
DR N-PSDB; AAK53401.

PT Nucleic acids encoding polypeptides with cytokine-like activities,  
PT useful in diagnosis and gene therapy -

PS Claim 20; Page 465-466; 6221pp; English.

CC The invention relates to polynucleotides (AAK51456-AAK53435) and the  
CC encoded polypeptides (AAM78323-AAM80302) that exhibit activity elating to  
CC cytokine, cell proliferation or cell differentiation or which may induce  
CC production of other cytokines in other cell populations. The  
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
CC peptide therapy. The polypeptides have various cytokine-like activities,  
CC e.g. stem cell growth factor activity, haematopoiesis regulating  
CC activity, tissue growth factor activity, immunomodulatory activity and  
CC activin/inhibin activity and may be useful in the diagnosis and/or  
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
CC inflammation.

CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666  
CC (AAM80020) are omitted as the relevant pages from the sequence listing  
CC were missing at the time of publication.

XX Sequence 2515 AA;

Query Match 11.0%; Score 87.5; DB 22; Length 2515;  
Best Local Similarity 24.2%; Pred. No. 40;  
Matches 38; Conservative 29; Mismatches 55; Indels 35; Gaps 6;

QY 5 VSELKPRVTVYTIQNGKMSSTVSEDFILPVYKGLGKGYQDGMWISGFEKQKAGY 64

DB 2352 IAEIQREPLVNSLNVNSGFTNEIEHSYKGEISGRKQNAARISGSHVEADP-- 2409

QY 65 VINLSKDTFKPVKKEIEKKKEENKPTFDVSKKK-----DNPQVNSQLNESH 113

DB 2410 -----KEYEE--BEHMPK---RKRKQYLSSEDEPDNPVLDRI-ETA 2449

QY 114 RKEDLQREHSQKSDSTKDTATVLDKNISKSSTN 150

DB 2450 QRCPEPTEPHATKEENSRDLIE--LPKTSSETNSTS 2484

RESULT 38

AAM80269

ID AAM80269 standard; Protein: 2515 AA.

XX AAM80269;

XX 06-NOV-2001 (first entry)

XX Human protein SEQ ID NO 3915.

XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;

KM vaccine; peptide therapy; stem cell growth factor; haematopoiesis;

KM tissue growth factor; immunomodulatory; cancer; leukaemia;

XX nervous system disorder; arthritis; inflammation.

XX Homo sapiens.

XX MO200157190-A2.

XX 09-AUG-2001.

XX 05-FEB-2001; 2001WO-US04098.

XX 03-FEB-2000; 2000US-0496914.

XX 27-APR-2000; 2000US-0560875.

XX 20-JUN-2000; 2000US-0598075.

XX 19-JUL-2000; 2000US-0620325.

```

XX WO200283876-A2.
XX
XX PD 24-OCT-2002.
XX
XX PE 27-MAR-2002; 2002WO-US09921.
XX
XX PR 29-MAR-2001; 2001US-280067P.
XX
XX PR 29-MAR-2001; 2001US-280068P.
XX
XX PR 16-MAY-2001; 2001US-291280P.
XX
XX PR 17-MAY-2001; 2001US-291829P.
XX
XX PR 17-MAY-2001; 2001US-291849P.
XX
XX PR 19-JUN-2001; 2001US-299428P.
XX
XX PR 20-JUN-2001; 2001US-29976P.
XX
XX PR 20-JUN-2001; 2001US-300001P.
XX
XX PA (INCY-) INCYTE GENOMICS INC.
XX
XX PI Daffo A, Jones AL, Tran AB, Dahl CR, Gietzen D, Chinn J;
XX
XX PI Dutour GE, Hillman JL, Yu JY, Tuason O, Yap PE, Amehy SR;
XX
XX PI Daugherty SC, Dam TC, Liu TF, Nguyen DA, Kleeefeld Y, Gerstein EH;
XX
XX PI Peralta CH, David MH, Lewis SA, Chen AJ, Paner SR, Harris B;
XX
XX PI Flores V, Marwaha R, Lo A, Lan KY, Uraehka ME;
XX
XX DR MPI; 2003-075543/07.
XX
XX DR N-PSDB; ABZ36315.
XX
XX PT New human secretory proteins and polynucleotides, useful for
XX
XX PT diagnosing, treating or preventing autoimmune/inflammatory disorders
XX
XX PT (e.g. AIDS), neurological disorders (e.g. Alzheimer's), or cell
XX
XX PT proliferations or cancers -
XX
XX PS Claim 27; SEQ ID NO 1057; 458bp + Sequence Listing; English.
XX
XX CC The invention relates to a secretory polynucleotide (designated sptm)
XX
XX CC comprising any of 567 polynucleotide sequences (ABZ35837-ABZ36403), a
XX
XX CC naturally occurring polynucleotide sequence at least 90 % identical to
XX
XX CC the polynucleotide sequence, a polynucleotide complementary to them or an
XX
XX CC RNA equivalent of them. The polypeptide or polynucleotide are useful for
XX
XX CC treating, preventing or diagnosing a disease or condition associated with
XX
XX CC the expression of functional SPTM. These are particularly useful for
XX
XX CC diagnosing, treating or preventing autoimmune/inflammatory disorders
XX
XX CC (e.g. acquired immunodeficiency syndrome, anaemia, asthma or Crohn's
XX
XX CC disease), neurological disorders (e.g. epilepsy, Huntington's disease,
XX
XX CC dementia, stroke, Alzheimer's disease, Creutzfeldt-Jakob disease,
XX
XX CC multiple sclerosis, cerebral palsy, Parkinson's disease, anxiety,
XX
XX CC schizophrenia or amnesia), or cell proliferative disorders (e.g.
XX
XX CC psoriasis, polycythemia vera, or cancers including adenocarcinoma,
XX
XX CC leukaemia, lymphoma, melanoma, myeloma, sarcoma or cancers of the brain,
XX
XX CC breast, cervix or prostate). The present sequence is one of the SPTM
XX
XX CC proteins of the invention (ABP75184-ABP75962).
XX
XX CC Note: The sequence data for this patent did not form part of the printed
XX
XX CC specification, but was obtained in electronic format directly from WIPO at
XX
XX CC ftp.wipo.int/pub/published_pct_sequences.
XX
XX SQ Sequence 3021 AA;
XX
XX Query Match 11.0%; Score 87.5; DB 24; Length 3021;
XX
XX Best Local Similarity 25.4%; Pred. No. 52;
XX
XX Matches 43; Conservative % 22; Mismatches 59; Indels 45; Gaps 7;
XX
XX QY 5 USELKHRYVTTTONGKEMSTIVSEDFILPVYKGLNGYQPCDGBWISFBEKKOAGY 64
XX
XX DB IAELOREPLLVNESIYNVSIGFTNIEIHSESYNKGELSSGRKNAAIBISHVEADP - 2728
XX
XX QY 65 VINSKOTFIKVPCKLEBKEEENKPLTFVDISKK-----DNPOVNHSQLNESH 113
XX
XX DB 2729 -----KEYVB--EEHMK--RRKKHYLTSEDDPDNDPNVLDSRIETAQ 2769
XX
XX QY 114 R-----KEDLR--EEHSOKSDSTKVDTATVL-DKNNISSKSTT 149
XX
XX DB 2770 RQCEBTPEPHDKENSRLDELPTTSSETJNSTSTRVWEKDEYSSEFTT 2818

```



RESULT 40  
AAC47057  
ID AAC47057 standard; Protein: 313 AA.  
XX  
AC AAC47057;  
XX  
DT 18-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 59270.  
XX  
KW Protein identification; signal transduction pathway; metabolic pathway;  
KM hybridisation assay; genetic mapping; gene expression control; promoter;  
XX termination sequence.  
XX  
OS Arabidopsis thaliana.  
XX  
EN EP1033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-0301439.  
XX  
PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.

PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 24-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.

PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 26-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 10.9%; Score 87; DB 21; Length 313;

Best Local Similarity 25.4%; Pred. No. 2.5; Mismatches 46; Indels 34; Gaps 6;

Matches 36; Conservative 26; Mismatches 46; Indels 34; Gaps 6;

QY 6 SBLKPHRVVTTIONGKEMSTIVSEBDPILPVYKGLKNGYOPDGWEISG-FEGKIDAGY 64  
DB 55 SSKPRNR--KTIQKGEIKSS-----PADG-KLSGKMKRKRKVG 91  
QY 65 VINLSKDTIKVFKKIEKGEENKPTPDVSKKQNPQVNSQLNESHKEDLQREHS 124  
DB 92 NVDISPEILKAIETKVKKKGNK---TKKKRAEBITRSSV-----EDLKRESKF 141  
QY 125 QKSDSTKVATVLDKQNISSK 146  
DB 142 KXSNKKKKQDMTSKKENKIEE 163

Search completed: February 10, 2004, 10:53:55  
Job time : 37.1775 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 10, 2004, 10:48:44 ; Search time 7.62413 Seconds  
(without alignments)  
1011.574 Million cell updates/sec

Title: US-10-067-385-8\_COPY\_610\_773

Perfect score: 848  
Sequence: 1 TTVEKFLNKGTGSEVSELRP.....ATVLAKNNISSKSTNNPNK 164

Scoring cable: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_41.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	97.5	11.5	1345	1 YH00 YEAST	P38800 saccharomyc
2	97.5	11.5	2468	1 MABP_HUMAN	P46821 homo sapien
3	95	11.2	348	1 CYL2_HUMAN	Q14093 homo sapien
4	95	11.2	1398	1 HSP2_PLAFA	P41001 plasmodium
5	93.5	11.0	649	1 HSP70_PYRSA	P37899 pyrenomonas
6	93	11.0	2464	1 MABP_MOUSE	P14873 mus musculu
7	92.5	10.9	1888	1 YDT2_SCHPO	Q14207 schizosacch
8	92	10.8	1202	1 DPMO_ASCIM	P22374 ascobolus 1
9	92	10.8	1875	1 MLPI_YEAST	P00245 saccharomyc
10	91	10.7	893	1 GYRA_STREP	P54112 staphylococ
11	90.5	10.7	443	1 DACH_BACSU	P08750 bacillus su
12	90.5	10.7	1332	1 SPT7_YEAST	P35177 saccharomyc
13	90	10.6	688	1 LIP_STAP	Q02510 staphylococ
14	89.5	10.6	433	1 TIG_STAM	Q09116 staphylococ
15	87.5	10.3	655	1 SKM1_YEAST	Q12469 saccharomyc
16	87.5	10.3	1702	1 IGAL2_HAEN	P45364 haemophilus
17	87	10.3	778	1 YFK8_YEAST	P43610 saccharomyc
18	87	10.3	6632	1 YFK8_CABEL	Q01761 caenorhabdi
19	86.5	10.2	406	1 NSB1_MOUSE	Q91135 mus musculu
20	86.5	10.2	1006	1 RAT1_YEAST	Q02792 saccharomyc
21	86	10.1	443	1 GLNA_PYRKO	Q08476 pyrococcus
22	86	10.1	914	1 PBPA_BACSU	P39793 bacillus su
23	86	10.1	5596	1 MDNI_HUMAN	Q9nu22 homo sapien
24	85.5	10.1	752	1 DRS1_YEAST	P32892 saccharomyc
25	85.5	10.1	1007	1 RGAL_YEAST	P39083 saccharomyc
26	85.5	10.1	2459	1 MABP_PAT	P15205 rattus norv
27	85	10.0	719	1 YBLE_STAP	Q10342 schizosacch
28	85	10.0	720	1 IF2_STAP	Q8cer4 staphylococ
29	84.5	10.0	657	1 Y040_MYCPN	P75062 mycoplasma
30	84.5	10.0	1694	1 IGA0_HAEN	P44969 haemophilus
31	84	9.9	510	1 HGV2_HAHO	Q02508 haemophilia
32	84	9.9	1251	1 RBP2_PLAAB	Q00799 plasmodium
33	83.5	9.8	360	1 FENR_PEA	P10933 plasmodium

34	83.5	9.8	479	1 U2R1_HUMAN	Q15695 homo sapien
35	83.5	9.8	678	1 GARP_PLAFA	P13816 plasmodium
36	83.5	9.8	731	1 HS9A_HUMAN	P07900 homo sapien
37	83.5	9.8	943	1 ARS2_DROME	Q9V9K7 drosophila
38	83	9.8	258	1 IF31_HUMAN	Q75822 homo sapien
39	83	9.8	270	1 TONB_HAEN	P42872 haemophilus
40	83	9.8	715	1 ADSV_HUMAN	Q95633 homo sapien
41	83	9.8	847	1 DNL4_CABEL	Q27474 caenorhabdi
42	83	9.8	3418	1 BRC2_HUMAN	P51587 homo sapien
43	82.5	9.7	439	1 GLNA_PYRAB	Q9uy99 pyrococcus
44	82.5	9.7	472	1 6PBD_PYRAB	Q9ch06 lactococcus
45	82.5	9.7	1553	1 TP2A_CHICK	Q62130 gallus gall

## ALIGNMENTS

RESULT 1	ID	YH00 YEAST	STANDARD;	PRT;	1345 AA.
AC	P38800;				
DT	01-FEB-1995 (Rel. 31, Last sequence update)				
DT	01-NOV-1995 (Rel. 32, Last annotation update)				
DE	Hypothetical 149.7 kDa protein in IRE1-KSPl intergenic region.				
GN	YHR080C.				
OS	Saccharomyces cerevisiae (Baker's yeast).				
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;				
OC	Saccharomycetales; Saccharomycetaceae; Saccharomyces.				
OX	NCBI_TaxID=4932;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=5288C / AB972;				
RX	MEDLINE=94378003; PubMed=8091229;				
RA	Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Dover J.,				
RA	De Z., Favell A., Fulton L., Gatliff S., Geisel C., Kirsten J.,				
RA	Kucaba T., Hillier L., Jier M., Johnston L., Langston Y.,				
RA	Latreille P., Louis E.J., Macri C., Mardis E., Meneses S., Mouser L.,				
RA	Nhan M., Rifkin L., Riles L., St Peter H., Trevaakie E., Vaughan R.,				
RA	Vignati D., Wilcox L., Wohlman P., Waterston R., Wilson R.,				
RA	Vandini M.,				
RT	"Complete nucleotide sequence of Saccharomyces cerevisiae chromosome				
RT	VIII."				
RL	Science 265:2077-2082(1994).				
CC	-1- SIMILARITY: TO YEAST YFL042C.				
CC					
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>				
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).				
CC					
DR	EMBL; U0556; AB68895.1; -.				
DR	PIR; S46817; S46817.				
DR	SGD; S0001122; YHR080C.				
DR	InterPro; IPR004182; GRAM_dom.				
DR	Pfam; PF02893; GRAM_1.				
DR	SMART; SM00568; GRAM_1.				
KW	Hypothetical protein; Transmembrane.				
FT	TRANSMEM 1198 1218 POTENTIAL.				
SQ	SEQUENCE 1345 AA; 149679 MW; 2FPAB94A66564C2 CRC64;				
Query Match	11.5%; Score 97.5; DB 1; Length 1345;				
Best Local Similarity	27.0%; Pred. No. 11;				
Matches	33; Conservative 18; Mismatches 38; Indels 33; Gaps 6;				
Qy	49 KGELEKYPQDGEISGFEKK-DAGVINLSKDTFVKPVFKKLEKKENKPTFDVSK 107				
Db	1109 KGALEK-----SVGQKVSVDYMLSELRDI-----SKAKSKPKVKVMK 1149				
Qy	108 KKNPNVNSQLNESHKEDLQREHSHQSKSDSTKDTATATVD--KNNISSKSTN--NP 162				

Db 1150 SHDKRPFPSKYE-----QKSSSRKSDDKDILTHLDPVQNNFSSEIRFMNKLSP 1201

Qy 163 NK 164

Db 1202 QK 1203

RESULT 2

MAPB\_HUMAN STANDARD; PRT; 2468 AA.

AC P46821; 01-NOV-1995 (Rel. 32, Created)

DT 01-NOV-1995 (Rel. 32, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Microtubule-associated protein 1B (MAP 1B) [contains: MAP1 light chain LC1].

GN MAP1B.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

OC NCBI\_Taxid=9606;

OK NCBI\_Taxid=9606;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Fetal brain;

RX MEDLINE=95104835; PubMed=7806212;

RA Lien L.V., Peener C., Fischbach N., Kunkel L.M.;

RT "Cloning of human microtubule-associated protein 1B and the identification of a related gene on chromosome 15.";

RL Genomics 22:273-280(1994).

CC -1- FUNCTION: THE FUNCTION OF BRAIN MAPS IS ESSENTIALLY UNKNOWN. PHOSPHORYLATED MAP1B MAY PLAY A ROLE IN THE CYTOSKELETAL CHANGES THAT ACCOMPANY NEURITE EXTENSION. POSSIBLY MAP1B BINDS TO AT LEAST TWO TUBULIN SUBUNITS IN THE POLYMER, AND THIS BRIDGING OF SUBUNITS MIGHT BE INVOLVED IN NUCLEATING MICROTUBULE POLYMERIZATION AND IN STABILIZING MICROTUBULES.

CC -1- SUBUNIT: 3 DIFFERENT LIGHT CHAINS, LC1, LC2 AND LC3, CAN ASSOCIATE WITH MAP1A AND MAP1B PROTEINS.

CC -1- DOMAIN: Has a highly basic region with many copies of the sequence KKEE and KKEI/V, repeated but not at fixed intervals, which is responsible for the binding of MAP1B to microtubules.

CC -1- PTM: LC1 is coexpressed with MAP1B. It is a polypeptide generated from MAP1B by proteolytic processing. It is free to associate with both MAP1A and MAP1B. It interacts with the amino-terminal region of MAP1B (by similarity).

CC -1- SIMILARITY: TO MAP1A.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

CC EMBL; L06237; AAA18904.1; -

CC DR Genew; HGNC:6836; MAP1B.

CC MIM; 157129; -

CC DR GO; GO:0005875; C:Microtubule associated complex; TMS.

CC DR InterPro; IPR000102; MAP1B\_neuraxin.

CC DR Pfam; PF00414; MAP1B\_neuraxin; 10.

CC DR PROSITE; PS00230; MAP1B\_NEURAXIN; 6.

CC KW Microtubules; Repeat; Phosphorylation.

CC FT CHAIN ? 2468 MAP1 LIGHT CHAIN LC1.

CC FT REPEAT 1878 1894 MAP1B 1.

CC FT REPEAT 1895 1911 MAP1B 2.

CC FT REPEAT 1912 1928 MAP1B 3.

CC FT REPEAT 1929 1945 MAP1B 4.

CC FT REPEAT 1946 1962 MAP1B 5.

CC FT REPEAT 1963 1979 MAP1B 6.

CC FT REPEAT 1997 2013 MAP1B 7.

CC FT REPEAT 2014 2030 MAP1B 8.

CC FT REPEAT 2031 2047 MAP1B 9.

FT REPEAT 2048 2064 MAP1B 10.

FT DOMAIN 589 790 LYS-RICH (HIGHLY BASIC, CONTAINS MANY KKEE AND KKEI/V REPEATS).

FT SQ SEQUENCE 2468 AA; 270618 MW; 540839CDBF09D461 CRC64;

Query Match

Best Local Similarity 24.8%; Pred. No. 21;

Matches 36; Conservative 29; Mismatches 47; Indels 33; Gaps 6;

Qy 4 KEPIANKDTGEVSELKPHRVTTIONGKMSSTIVSEEDTILVYKGELEKGYQPOCMEI 63

Db 584 EKVMVKDKPVKTEKTPSVTEKEVPSEKPS-----PV-XAEVA-----EK 623

Qy 64 SGFEGKDKAGYVYNLSKDTFLKPVFKKIEBKKEBKPPYDVGKKKDNPPVNHSQLNESH 123

Db 624 QATDVAPKAAKKEKTKKTKV-----BDKKEBKPKKXAKVDKTPI---KKEKP 675

Qy 124 KKEDLQRE-----EHSQKSDSTVDY 143

Db 676 KKEVVKKEVKEIKKEKKEPKKXEV 700

RESULT 3

CYL2\_HUMAN STANDARD; PRT; 348 AA.

ID CYL2\_HUMAN

AC 014053;

DT 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DB Cylidin II (multiple-band polypeptide II).

GN CYLC2 OR CYL2.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

OC NCBI\_Taxid=9606;

OK NCBI\_Taxid=9606;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Testis;

RX MEDLINE=95255491; PubMed=7737358;

RA Hess H., Heid H., Zimbelmann R., Franke W.W.;

RT "The protein complexity of the cytoskeleton of bovine and human sperm heads: the identification and characterization of cylidin II.";

RL Exp. Cell Res. 218:174-182(1995).

CC -1- FUNCTION: POSSIBLE ARCHITECTURAL ROLE DURING SPERMATOGENESIS. MAY BE INVOLVED IN SPERMATID DIFFERENTIATION.

CC -1- SUBCELLULAR LOCATION: CALYX; SPERM HEAD CYTOSKELETAL STRUCTURE.

CC -1- TISSUE SPECIFICITY: Testis.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

CC EMBL; Z46788; CAA6752.1; -

CC DR PIR; I37271; I37271.

CC DR Genew; HGNC:2583; CYLC2.

CC MIM; 604035; -

CC DR GO; GO:0005200; F:Structural constituent of cytoskeleton; TMS.

CC DR Cytoskeleton; Structural protein; Repeat; Sperm; Spermatogenesis.

CC FT DOMAIN 25 347 31 X 3 AA REPEATS OF K-K-X.

CC FT DOMAIN 157 240 3 X APPROXIMATE TANDEM REPEATS.

CC FT REPEAT 157 184 1.

CC FT REPEAT 185 212 2.

CC FT REPEAT 213 240 3.

CC SQ SEQUENCE 348 AA; 39079 MW; D86765599C1809E7 CRC64;

Query Match

Best Local Similarity 30.9%; Pred. No. 3; 39; Indels 26; Gaps 7;

Matches 38; Conservative 20; Mismatches 39; Indels 26; Gaps 7;

```

OY      49  NGELAEKGYQFPGWMEISGFEKKDAGVIVLNSDCTIKKVPFKLEKKKEBKEMPT'---DV 105
DB      205  ESEEGK-----GTEDSDKKKKDS-----KKGDSAIELQAVQADKKEDCKDANKADE 256
OY      106  SK--KKDPOVNHSQLN-----ESHKPDLOREHSGKSDSTD---VTATVLDKNNI 153
DB      257  SKDAGKDKMKELKKGGKDKKKKGSSTDSKDVKKE--SKDQATDAAKKVAKKQTEKESA 313
OY      154  SSK 156
DB      314  DSK 316

RESULT 4
TOP2_PLAFK
ID      TOP2_PLAFK      STANDARD;      PRT; 1398 AA.
AC      P41001.
DT      01-FEB-1995 (Rel. 31, Created)
DT      01-FEB-1995 (Rel. 31, Last sequence update)
DT      30-MAY-2000 (Rel. 39, Last annotation update)
DE      DNA topoisomerase II (EC 5.99.1.3).
GN      TOP2.
OS      Plasmodium falciparum (isolate KI / Thailand).
OC      Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.
OX      NCBI_TaxID=5839;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=94316496; PubMed=8041616;
RA      Cheesman S., McAliese S., Goman M., Johnson D., Horrocks P.,
RA      Ridley R.G., Kilbey B.J.;
RT      "The gene encoding topoisomerase II from Plasmodium falciparum.";
RL      Nucleic Acids Res. 22:2547-2551(1994).
CC      -1- FUNCTION: CONTROL OF TOPOLOGICAL STATES OF DNA BY TRANSIENT
CC      BREAKAGE AND SUBSEQUENT REJOINING OF DNA STRANDS. TOPOISOMERASE II
CC      MAKES DOUBLE-STRAND BREAKS.
CC      -1- CATALYTIC ACTIVITY: ATP-dependent breakage, passage and rejoining
CC      of double-stranded DNA.
CC      -1- SUBUNIT: Homodimer (By similarity).
CC      -1- SUBCELLULAR LOCATION: Nuclear.
CC      -1- MISCELLANEOUS: EUKARYOTIC TOPOISOMERASE I AND II CAN RELAX BOTH
CC      NEGATIVE AND POSITIVE SUPERCOILS, WHEREAS PROKARYOTIC ENZYMES
CC      RELAX ONLY NEGATIVE SUPERCOILS.
CC      -1- SIMILARITY: Belongs to the type II topoisomerase family.
CC
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC
CC      -----
DB      EMBL; X79345; -; NOT_ANNOTATED_CDS.
DB      HSSP; P06786; 1BGM.
DB      InterPro; IPR003594; ATPbind_ATPase.
DB      InterPro; IPR003957; CBPA_NFYD_topis.
DB      InterPro; IPR001241; DNA_topoisoi.
DB      InterPro; IPR002205; DNA_topoisoiV.
DB      Pfam; PF00204; DNA_gyraseb; 1.
DB      Pfam; PF00521; DNA_topoisoiV; 1.
DB      Pfam; PF02518; HATPase_c; 1.
DB      PRINTS; PR00615; CCAATSUBUNTA.
DB      PRINTS; PR00418; TP12FAMILY.
DB      PRODOM; PD000742; DNA_topoisoiV; 1.
DB      SMART; SM00387; HATPase_c; 1.
DB      SMART; SM00433; TOP2c; 1.
DB      SMART; SM00434; TOP4c; 1.
DB      PROSITE; PS00177; TOPOISOMERASE_II; 1.
DB      K1 isomerase; TOPOISOMERASE; DNA-binding; ATP-binding; Nuclear protein.
FT      NP_BIND 144 149      ATP (POTENTIAL).
FT      ACT_SITE 830 830      DNA CLEAVAGE (BY SIMILARITY).
FT      DOMAIN 271 281      POLY-ASN.
FT      DOMAIN 308 316      POLY-ASN.

```

```

FT DOMAIN 1089 1093 POLY-LYS.
FT DOMAIN 1227 1234 POLY-LYS.
SQ SEQUENCE 1398 AA; 161029 MW; 6AAD7BE88FE5EB9 CRC64;

Query Match 11.2%; Score 95; DB 1; Length 1398;
Best Local Similarity 23.8%; Pred. No. 17;
Matches 36; Conservative 36; Mismatches 53; Indels 26; Gaps 6;

OY 29 NGKESSTIVEEEDFIL--PYKNGELEKGYQFDGMEISGFGKKDAGYVNLKSDTFIKP 86
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
DB 1133 NEELIAGTTVDKDYLLSMFPFSITLEK---VEOLLTQLKEKEBELIRIYETMTWK 1189
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
OY 87 VFKKLEE-----KKEENKPTFDVSKKDDNPQVNHSSQLNESHKEDLQREHSQ 135
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
DB 1190 DLEKVEELIERQRAVELSNRESNK--PKAAKQ-----GFSNMKKKKKKKSSDESE 1242
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
OY 136 ---KSDSTQVATATVLDKNNISSKSTNNPN 163
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
DB 1243 GGDTSDSSEFLVNTLNKKVNTKTTTSSNN 1273
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |

RESULT 5
HS70 PYRSA
ID HS70 PYRSA STANDARD; PRT; 649 AA.
AC P37899;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Heat shock 70 kDa protein.
GN HSP70.
OS Pyrenomonas salina.
OC Nucleomorph.
CC Eukaryota; Cryptophyta; Cryptomonadaceae; Pyrenomonas.
OX NCBI_TaxID=3034;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94268506; Pubmed=8208251;
RA Hofmann C.J.B., Rensing S.A., Haeuber M.M., Martin W.F., Mueller S.B.,
RA Couch D., McFadden G.I., Iglot G.L., Møller U.-G.;
RT "The smallest known eukaryotic genomes encode a protein gene: towards
RT an understanding of nucleomorph functions.";
RL Mol. Gen. Genet. 243:600-604(1994).
CC -I- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 70 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-ib.ch).
CC -----
CC EMBL; X72621; CAA51197.1; -.
CC DR PIR; S42488; S42488.
CC DR HSSP; P08109; ICKR.
CC DR InterPro; IPR001023; Hsp70.
CC DR Pfam; PF00012; HSP70_1.
CC DR PRINTS; PR00301; HEATSHOCK70.
CC DR ProDom; PD000089; HSP70_1.
CC DR PROSITE; PS00297; HSP70_1.
CC DR PROSITE; PS00329; HSP70_2.
CC DR PROSITE; PS01036; HSP70_3.
CC DR ATP-binding; Heat shock; Nucleomorph.
CC KW SEQUENCE 649 AA; B627B08BF90C9164 CRC64;
CC -----

Query Match 11.0%; Score 93.5; DB 1; Length 649;
Best Local Similarity 24.5%; Pred. No. 97;
Matches 38; Conservative 26; Mismatches 50; Indels 41; Gaps 7;

OY 7 ILNKDNGEVELKPHRTVTLONG---KEMSSRTVSEBDPLLVYNGELEKGYQFQSMEL 63
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
DB 491 ILNWSADKSTGKSNKTTINDKGRSLKEIRVEVEAR---KYKTEDEK----- 537

```

QY 64 SGEFGKDA-----GVINLS---KDTFIKPVKIEKKKEENKPTFVSKKKNPQVN 115  
 DB 538 --LDKKLEAKNSLENTAYNIRNTRD-----EKLEKXIQEDKXSIKBEKVE----- 582  
 QY 116 HSQLNESHKEDLQREHSOKSDSTQVATVLDK 150  
 DB 583 --VLEFIETNEDEKKEEYERKEKELKNFANPIISK 615

RESULT 6  
 MAPB\_MOUSE STANDARD; PRT; 2464 AA.  
 ID MAPB\_MOUSE  
 AC P14873;  
 DT 01-APR-1990 (Rel. 14, Created)  
 DT 01-APR-1990 (Rel. 14, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Microtubule-associated protein 1B (MAP 1B) (MAP1.2) (MAP1(X))  
 DE [Contains: MAP1 light chain LC1].  
 GN MAP1B OR MTAP1B OR MTAP5.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxId=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND DOMAIN.  
 RX STRAIN=Swiss Webster; TISSUE=Brain;  
 RX MEDLINE=90094539; PubMed=2480963;  
 RA Noble M., Lewis S.A., Cowan N.J.;  
 RT "The microtubule binding domain of microtubule-associated protein  
 MAP1B contains a repeated sequence motif unrelated to that of MAP2  
 and tau";  
 RL J. Cell Biol. 109:3367-3376(1989).  
 CC -1- FUNCTION: THE FUNCTION OF BRAIN MAPS IS ESSENTIALLY UNKNOWN.  
 CC PHOSPHORYLATED MAP1B MAY PLAY A ROLE IN THE CYTOSKELETAL CHANGES  
 CC THAT ACCOMPANY NEURITE EXTENSION. POSSIBLY MAP1B BINDS TO AT LEAST  
 CC TWO TUBULIN SUBUNITS IN THE POLYMER, AND THIS BRIDGING OF SUBUNITS  
 CC MIGHT BE INVOLVED IN NUCLEATING MICROTUBULE POLYMERIZATION AND IN  
 CC STABILIZING MICROTUBULES.  
 CC -1- SUBUNIT: 3 DIFFERENT LIGHT CHAINS, LC1, LC2 AND LC3, CAN ASSOCIATE  
 CC WITH MAP1A AND MAP1B PROTEINS.  
 CC -1- DOMAIN: Has a highly basic region with many copies of the sequence  
 CC KKEE and KKEI/V, repeated but not at fixed intervals, which is  
 CC responsible for the binding of MAP1B to microtubules.  
 CC -1- PTM: LC1 IS COEXRESSED WITH MAP1B. IT IS A POLYPEPTIDE GENERATED  
 CC FROM MAP1B BY PROTEOLYTIC PROCESSING. IT IS FREE TO ASSOCIATE WITH  
 CC BOTH MAP1A AND MAP1B. IT INTERACTS WITH THE AMINO-TERMINAL REGION  
 CC OF MAP1B.  
 CC -1- SIMILARITY: TO MAP1A.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 CC EMBL; X51396; CAA35761.1; -  
 CC DR PIR; S07549; GRMSP1.  
 CC DR MGI; MGI:1306778; Mtap1b.  
 CC DR GO; GO:0016358; P: dendritic morphogenesis, IMP.  
 CC DR GO; GO:0001578; P: microtubule bundling, IMP.  
 CC DR InterPro; IPR00102; MAP1B neuraxin.  
 CC DR Pfam; PF00414; MAP1B neuraxin; 10.  
 CC DR PROSITE; PS00230; MAP1B NEURAXIN; 7.  
 CC Microtubules; Repeat; Phosphorylation.  
 CC CHAIN ? 2464 MAP1 LIGHT CHAIN LC1.  
 FT REPEAT 1874 1890 MAP1B 1.  
 FT REPEAT 1891 1907 MAP1B 2.  
 FT REPEAT 1908 1924 MAP1B 3.  
 FT REPEAT 1925 1941 MAP1B 4.  
 FT REPEAT 1942 1958 MAP1B 5.  
 FT REPEAT 1959 1975 MAP1B 6.

FT REPEAT 1993 2009 MAP1B 7.  
 FT REPEAT 2010 2026 MAP1B 8.  
 FT REPEAT 2027 2043 MAP1B 9.  
 FT REPEAT 2044 2060 MAP1B 10.  
 FT DOMAIN 589 787 LYS-RICH (HIGHLY BASIC, CONTAINS MANY  
 FT KKEE AND KKEI/V REPEATS)  
 SQ SEQUENCE 2464 AA; 270408 MW; FBD3DD99CFDDBA87 CXC64;  
 Query Match 11.0%; Score 93; DB 1; Length 2464;  
 Best Local Similarity 27.8%; Pred. No. 45;  
 Matches 49; Conservative 24; Mismatches 57; Indels 46; Gaps 11;  
 QY 4 KEFLINKQGEVSELSKPHRVTTIIONGKMSSTIVSEEDFILPVYKGLKGYQDFGWEI 63  
 DB 584 EKVLYKKQKPVKTESK---SVT---KKEVSS---KEQ--SPV-KAEVA-----EK 623  
 QY 64 SGEFGKDAGYINISKDTFIKPVFK-KIEKKKEENKPTFVSKKKNPQVNSQINES 122  
 DB 624 QATESKP-----KTVKDKVKKKEIKTKLEKKEE--KKEVKKEDKTPPL--KKDEK 672

QY 123 HKREDLQRE-----EHSQKSDSTQVATVLDKNNISKSTNNPK 164  
 DB 673 PKKEVKKKEIKKKEIKKEKKEKELKKEVKETPLKADAKKEVKKKEKKEKKEPKK 728

RESULT 7  
 YDIT2 SCHPO STANDARD; PRT; 1888 AA.  
 ID YDIT2 SCHPO  
 AC 014207;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Hypothetical protein C6B12.02c in chromosome I.  
 GN SPAC6B12.02C.  
 OS Schizosaccharomyces pombe (fission yeast).  
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
 OC Schizosaccharomycetiales; Schizosaccharomycetaceae;  
 OC Schizosaccharomyces.  
 OX NCBI\_TaxId=4896;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX STRAIN=972;  
 RX MEDLINE=21848401; PubMed=11859360;  
 RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,  
 RA Sgourou J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,  
 RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,  
 RA Collins M., Connor R., Cronin A., Davis P., Fellwell T., Fraser A.,  
 RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,  
 RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagsels K.,  
 RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,  
 RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,  
 RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitch E.,  
 RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,  
 RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,  
 RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,  
 RA Woodward J., Volckaert G., Aert R., Robben J., Grynoprez B.,  
 RA Weljens I., Vanetseis E., Rieger M., Schaefer M., Mueller-Auer S.,  
 RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,  
 RA Borzym K., Langer I., Beck A., Lehnach H., Reinhardt R., Pohl T.M.,  
 RA Eger P., Zimmermann W., Wedler R., Wambutt R., Purnelle B.,  
 RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,  
 RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,  
 RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garcon A., Thode G.,  
 RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,  
 RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,  
 RA Cerruti L., Lowe T., McCombie M.R., Paulsen I., Potashkin J.,  
 RA Shpakovski G.V., Useery D., Barrell B.G., Nurse P.;  
 RT Nature 415:871-880(2002).  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (potential).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL\_298531; CAB11064.1; -  
 DR PIR\_139009; T39009.  
 DR GenedB\_Spomb; SPAC6B12.02c; -  
 KW Hypothetical protein; Transmembrane.  
 FT TRANSMEM 697 717 POTENTIAL.  
 FT TRANSMEM 755 775 POTENTIAL.  
 FT TRANSMEM 866 886 POTENTIAL.  
 FT TRANSMEM 915 935 POTENTIAL.  
 FT TRANSMEM 977 997 POTENTIAL.  
 FT TRANSMEM 1033 1053 POTENTIAL.  
 FT TRANSMEM 1336 1356 POTENTIAL.  
 FT TRANSMEM 1645 1665 POTENTIAL.  
 FT DOMAIN 1662 1665 POLY-LEU.  
 SQ SEQUENCE 1888 AA; 217432 MW; 8AD3BCE32397C29 CRC64;

Query Match 10.8%; Score 92.5; DB 1; Length 1888;  
 Best Local Similarity 23.1%; Pred. No. 36;  
 Matches 39; Conservative 25; Mismatches 76; Indels 29; Gaps 5;

QY 1 TVVKEFLINKGVESELKPHRYTVLTQNGKEMSTIVE-BDFILPYKGELEKGYQD 59  
 DB 394 TSENPPOLVANAASTIPVYRTTKMKKRNKRYVEKLPDLIESY----- 442  
 QY 60 GWEISGFEGKDGAGYVNLKDTPIKVPFKIEEKEEENKPTFDVSKKQNPQVNSQL 119  
 DB 443 -----GKAPKPLRYVARSSNHP--KMTRRKQNDKKYFSDKSDKQVLDVLS 492  
 QY 120 NESHRKBDLQREBSHQSQSDS-TEDVATATVLDKN-----NISKSTTN 161  
 DB 493 DWYSGKHLELVQSHSYKRSKSVGAGNIFSVNKKSHSVINAKTAANN 541

RESULT 8  
 DROM\_ASCIM STANDARD; PRT; 1202 AA.  
 AC P22374;  
 DT 01-AUG-1991 (Rel. 19, Created)  
 DT 01-AUG-1991 (Rel. 19, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE Probable DNA polymerase (EC 2.7.7.7).  
 OS Ascobolus immerus.  
 OG Mitochondrion.  
 OC Plasmid pA12.  
 CC Baktaryota; Fungi; Ascomycota; Pezizomycotina; Pezizomycetes;  
 CC Pezizales; Ascobolaceae; Ascobolus.  
 OK NCBI\_TaxId=5191;  
 RN NCBI\_TaxId=5191;  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=2/1;  
 RX MEDLINE=90066356; PubMed=2573821;  
 RA Kempken F., Weinhardt F., Esser K.;  
 RT "In orfanello replication and viral affinity of linear,  
 RT extrachromosomal DNA of the ascomycete Ascobolus immerus.";  
 RL Mol. Genet. 218:523-530 (1989).  
 CC -1- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate = N diphosphate  
 CC + {DNA} (N).  
 CC -1- MISCELLANEOUS: THIS DNA POLYMERASE REQUIRES A PROTEIN AS A PRIMER  
 CC (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO DNA POLYMERASE TYPE-B FAMILY. STRONG TO  
 CC DNA POLYMERASE OF OTHER FUNGAL AND PLANTS MITOCHONDRIAL PLASMIDS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL\_X15982; CA34106.1; -  
 DR PIR\_S05362; S05362.  
 DR InterPro: IPR006172; DNA\_pol\_B.  
 DR InterPro: IPR004868; DNA\_pol\_B\_2.  
 DR Pfam: PF03175; DNA\_pol\_B\_2; 1.  
 DR SMART: SM00486; POLBc\_1.  
 DR PROSITE: PS00116; DNA\_POLYMERASE\_B\_1.  
 KW Transferrase; DNA-directed DNA polymerase; DNA replication;  
 KW DNA-binding; Plasmid; Mitochondrion.  
 SQ SEQUENCE 1202 AA; 138279 MW; 51D41FCEDBF2C0E CRC64;

Query Match 10.8%; Score 92; DB 1; Length 1202;  
 Best Local Similarity 22.5%; Pred. No. 24;  
 Matches 42; Conservative 37; Mismatches 76; Indels 32; Gaps 9;

QY 6 FILNK-----DYGVESLKPHRYTVLTQNGKEMSTI---VSEDDF--ILPYKGELEK-- 54  
 DB 311 FVYNAKIKPEPTGVNRSIGFAGNT-TLTDKTLTKTLALFLEREDITHVMSYDGDIDESK 369  
 QY 55 -----GYQPDGWEISGFEGKDGAGYVNLKDTPIKVPFKIE-----EKKEENK-PTF 103  
 DB 370 FPKGISLFDKPKLTKTEGTYKTYNTPPIKDDIVVKDINKINFGDLPTPTMDSKPNL 429  
 QY 104 DYSKKDNPQVNSHQLNSHRSKEDLQREBSHQSQSDSTKDTATATVLDKNNT-----SSK 156  
 DB 430 KLNKDKTSGEIRNTIKKNNQSYDI--IGHMTINDSNTTFRRAYDNSIIKLFYTDSD 487  
 QY 157 STTNPN 163  
 DB 488 GNTNDPN 494

RESULT 9  
 MLPI\_YEAST STANDARD; PRT; 1875 AA.  
 AC Q02455;  
 DT 01-OCT-1993 (Rel. 27, Created)  
 DT 01-JUN-1994 (Rel. 29, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Myosin-like protein MLPI.  
 GN MLPI OR YKR095W OR YKR415.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 CC Baktaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 CC Saccharomycetales; Saccharomycetaceae; Saccharomyces.  
 OK NCBI\_TaxId=4932;  
 RN NCBI\_TaxId=4932;  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=S288C;  
 RX MEDLINE=93247549; PubMed=8483450;  
 RA Koelling R., Nguyen T., Chen E.Y., Botstein D.;  
 RT "A new yeast gene with a myosin-like heptad repeat structure.";  
 RL Mol. Genet. 237:359-369 (1993).  
 CC -1- FUNCTION: MYOSIN-LIKE PROTEIN THAT IS PROBABLY INVOLVED IN DNA  
 CC REPAIR.  
 CC -1- SIMILARITY: SOME TO THE TPR ONCOGENE.  
 CC -1- CAUTION: REF. 2 MISQUOTES THE GENE NAME AS "MPL1".  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial

entires requires a license agreement (see <http://www.isb-sib.ch/announce>  
or send an email to [license@isb.sib.ch](mailto:license@isb.sib.ch)).  
-----  
CC EMBL; L01992; AAA34783.1; -;  
DR EMBL; X73541; CAAS1948.1; -;  
DR EMBL; Z28320; CAA82174.1; -;  
DR PIR; S38173; S38173.  
DR SGD; S0001803; MRP1.  
DR GO; GO:0005635; Cy:nuclear membrane; IDA.  
DR GO; GO:0006606; P:protein-nucleus import; IDA.  
DR GO; GO:0006606; P:protein-nucleus import; IDA.  
KW Coiled coil; DNA repair.  
FT DOMAIN 69 487 COILED COIL (POTENTIAL).  
FT DOMAIN 531 1678 COILED COIL (POTENTIAL).  
FT DOMAIN 1834 1868 COILED COIL (POTENTIAL).  
FT COMPLECT 301 301 R-> A (IN REF.1).  
SQ SEQUENCE 1875 AA; 218455 MW; 683AD0D3C9066867 CRC64;

Query Match 10.8%; Score 92; DB 1; Length 1875;  
Best Local Similarity 25.0%; Pred. No. 39;  
Matches 48; Conservative 36; Mismatches 62; Indels 46; Gaps 10

Dy 7 ILNKDTGVSELKPRVTVTTONGKEMASTYSEDF-----LPVKG--LEKYQF 58  
::|::|::|::|::|::|::|::|::|::|:  
Db 660 LINK---EIDLYDSKDSDISIKGEXKSRIIAERFKLLNSTLDLTAEENDQLKKRFDY 716  
::|::|::|::|::|::|::|::|::|::|:  
Cy 59 DGEWISGFEGEKDA-----GYVLNSKDPFIKEVPFKKIEBK-----KEBENK 100  
::|::|::|::|::|::|::|::|::|::|:  
Db 717 ---LQNITLKQDSKTHTETLVNYVCCKSLSVIETELNLAKRGQLRVHLEGNLKOFLNK 772  
::|::|::|::|::|::|::|::|::|::|:  
Cy 101 PTFDYSKKKDNPOVNHSQUNESHK-KEDE---QRSHSKSSTVDVTN----TVDLNK 151  
::|::|::|::|::|::|::|::|::|::|:  
Db 773 ----LSPKKDSRLRWITQILOKEREDILLKETRSCKRKIDDELDAISLKGETSYQDH 828  
::|::|::|::|::|::|::|::|::|::|:

Cy 152 NISSKSTNNPN 163  
::|::|::|::|::|::|::|::|::|::|:  
Db 829 HIKOLEEDNNSN 840  
::|::|::|::|::|::|::|::|::|::|:

RESULT 10  
CYRA\_STAEP STANDARD; PRT; 893 AA.

ID CYRA\_STAEP P54112;  
AC P54112;  
DT 01-OCT-1996 (Rel. 34, Created)  
DD 28-FEB-2003 (Rel. 41, Last sequence update)  
DI 28-FEB-2003 (Rel. 41, Last annotation update)  
DE DNA gyrase subunit A [EC 5.99.1.3].  
EN GYA OR SE0005.  
OS Staphylococcus epidermidis.  
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.  
OX NCBI\_TaxID=1282;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 12228.  
RA Zhang Y., Ren S., Li H., Fu G., Lu L., Lu G., Jia J., Tu Y., Qin Z.,  
Chen Z., Wen Y.;  
RU Submitted (NOV-2002) to the EMBL/genbank/DBJ databases.  
RX MEDLINE=99102204; PubMed=166207;  
RY SEQUENCE OF 1-94 FROM N.A.; AND NOTAGNESIS OF SER-84.  
SA Streedharan S., Peterson L.R., Fisher L.M.;  
SB "Citrobacter resistance in coagulase-positive and -negative  
staphylococci: role of mutations at serine 94 in the DNA gyrase A  
protein of Staphylococcus aureus and Staphylococcus epidermidis.";  
SC Antimicrob. Agents Chemother. 35:2151-2154(1991).  
SD FUNCTION: DNA GYASE NEGATIVELY SUPERCOLLS CLOSED CIRCUAR DOUBLE-  
STRANDED DNA IN AN ATP-DEPENDENT MANNER AND ALSO CATALYZES THE  
INTERCONVERSION OF OTHER TOPOLOGICAL ISOMERS OF DOUBLE-STRANDED  
DNA RINGS, INCLUDING CATENANES AND KNOTTED RINGS.  
SE CATALYTIC ACTIVITY: ATP-dependent breakage, passage and rejoining  
of double-stranded DNA.  
SF SUBUNIT: MADE UP OF TWO CHAINS. THE A CHAIN IS RESPONSIBLE FOR DNA  
BREAKAGE AND REJOINING; THE B CHAIN CATALYZES ATP HYDROLYSIS. THE

```
CC      ENZYME FORMS AN A2B2 TETRAMER.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC      The European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; AE016744; AA003602.1; -.
DR      EMBL; S72803; AB20672.1; -.
DR      PIR; A49832; A49832.
DR      HSSP; P09097; LAB4.
DR      InterPro; IPR002205; DNA_topoisolv.
DR      Pfam; PF03989; DNA_gyraseA_C_6.
DR      ProDom; PD000742; DNA_topoisolv; 1.
DR      SMART; SM00434; TOPAC; 1.
DR      TrIPAPAS; TRIGR01063; Gyra; 1.
KW      Topoisomerases; Isomerase; DNA-binding; Antibiotic resistance;
KM      Complete proteome.
FT      ACT_SITE 123 123          DNA CLEAVAGE (BY SIMILARITY).
FT      MUTAGEN  84  84           S->E: RESISTANT TO CIPROFLOXACIN.
SQ      SEQUENCE   893 AA; 100113 MW; 2A6A7CD345A526CE CRC64;
Query Match              10.7%; Score 91; DB 1; Length 893;
Best Local Similarity    24.9%; Pred. No. 21;
Matches 43; Conservative 31; Mismatches 75; Indels 24; Gaps 8;
QY      1 TTVAEPFLNDOTGEVSELKPHRYVTWTLONGKMS-STVSSEDPFLPVYKELEKGYPD 59
DB      727 TPVSEYRLSNRGGKIX-----TATITERNGNTVCITVTGEEDLMTVTVAGVI---IRLD 779
QY      60 GWESIFEGEKKNADY-VYNLSKDTFIKVPFKLIEKKKEEN-----KPTDVSKRK 109
DB      780 VHDISO-MGRRAQGCVRLMKGDGFSTVAKVNEDNEFNADAQGSTTFETADVSEV 838
QY      110 DNPGVNHSOLNESHREKEDLGREE-HSGQSDSTVDVATVLIDKNNISKSSTNN 161
DB      839 D----DQTPGAHTTBGDAEMESVESPENDRIDIRDQFMDRVVEDIESASN 887
RESULT 11
DACA_BACSU
ID     _DACA_BACSU        STANDARD;             PRT;         443 AA.
AC     P08750.
DT     01-AUG-1988 (Rel. 08, Created)
DD     01-OCT-1994 (Rel. 30, Last sequence update)
DT     28-FEB-2003 (Rel. 41, Last annotation update)
DE     d-alanyl-D-alanine carboxypeptidase precursor (EC 3.4.16.4) (DD-
DE     peptidase) (OD-carboxypeptidase) (CPase) (PBPs).
GN     DACA.
OS     Bacillus subtilis.
OC     Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
CX     NCBI_TaxID=1423;
RN     [1]
RP     SEQUENCE FROM N.A.
RC     STRAIN=168;
RX     MEDLINE=96051385; PubMed=7584024;
RA     Ogasawara N., Nakai S., Yoshikawa H.;
RT     "Systematic sequencing of the 180 kilobase region of the Bacillus
RT     subtilis chromosome containing the replication origin.";
RL     DNA Res. 1:1-14(1994).
RN     [2]
RP     SEQUENCE FROM N.A.
RC     STRAIN=168;
RX     MEDLINE=96044033; PubMed=9384377;
RA     Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA     Azavedo V., Bertero M.G., Besieres P., Bolotin A., Borchert S.,
RA     Borries R., Boursier L., Brans A., Brann M., Bridnell S.C., Bron S.,
RA     Brouillet S., Bruenchi C.V., Caldwell B., Capano V., Carter N.M.,
RA     Choi S.K., Codani U.J., Comerton I.F., Cummings M.J., Daniel R.A.,
RA     Denizot F., Devine K.M., Ducresthoft A., Ehlich S.D., Emerson P.T.,
```



RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,  
 RA Fritz C., Fujita M., Fujita Y., Fuma S., Gallizzi A., Galleron N.,  
 RA Guiseppe G., Guy B.J., Haga K., Haelech J., Harwood C.R., Henaut A.,  
 RA Hilbert H., Holsteppel S., Hosono S., Hullo M.F., Itaya M., Jones L.,  
 RA Joris B., Karamata D., Kasahara Y., Kleier-Bianchini M., Klein C.,  
 RA Kobayashi Y., Koester P., Koningslehn G., Krogg S., Kumano M.,  
 RA Kurita K., Lapides A., Lardinois S., Lauber J., Lazarevic V.,  
 RA Lee S.M., Levine A., Liu H., Masuda S., Manuel C., Medigue C.,  
 RA Medina N., Meliado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,  
 RA Moore D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,  
 RA Parro V., Pohl T.M., Portecelle D., Porwollik S., Prescott A.M.,  
 RA Presecan E., Puig P., Purnelle B., Rapoport G., Rey M., Reynolds S.,  
 RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,  
 RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,  
 RA Sekiguchi J., Sekowska A., Seror S.J., Seror P., Shin B.S., Soldo B.,  
 RA Sorokin A., Taccini E., Takagi T., Takahashi H., Takemaru K.,  
 RA Taneuchi M., Tamakoshi A., Tanaka T., Terstera P., Tognoni A.,  
 RA Tostato V., Uchiyama S., Vandenbol M., Vannier P., Vassarotti A.,  
 RA Viari A., Wambut R., Wedler E., Wedler H., Weitzenger T.,  
 RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,  
 RA Yoshida K., Yoshikawa H.F., Zumslein E., Yoshikawa H., Danchin A.,  
 RT "The complete genome sequence of the Gram-positive bacterium *Bacillus*  
 RT *subtilis*." ;  
 RL Nature 390:249-256(1997).  
 [3]  
 RP SEQUENCE OF 32-102.  
 RX MEDLINE=80182289; PubMed=6768745;  
 RA Waxman D.J., Strominger J.L.;  
 RT "Sequence of active site peptides from the penicillin-sensitive D-  
 RT alanine carboxypeptidase of *Bacillus subtilis*. Mechanism of  
 RT penicillin action and sequence homology to beta-lactamases." ;  
 RL J. Biol. Chem. 255:3964-3976(1980).  
 [4]  
 RP SEQUENCE OF 103-443 FROM N.A.  
 RX MEDLINE=86250602; PubMed=3087956;  
 RA Todd J.A., Roberts A.N., Johnstone K., Pigot P.J., Winter G.,  
 RA Ellar D.J.;  
 RT "Reduced heat resistance of mutant spores after cloning and  
 RT mutagenesis of the *Bacillus* subtilis gene encoding penicillin-binding  
 RT protein 5." ;  
 RL J. Bacteriol. 167:257-264(1986).  
 [5]  
 RP SEQUENCE OF 414-443.  
 RX MEDLINE=81117303; PubMed=6780559;  
 RA Waxman D.J., Strominger J.L.;  
 RT "Primary structure of the COOH-terminal membranous segment of a  
 RT penicillin-sensitive enzyme purified from two *Bacilli*." ;  
 RL J. Biol. Chem. 256:2067-2077(1981).  
 -1- FUNCTION: REMOVES C-TERMINAL D-ALANYL RESIDUES FROM SUGAR-PEPTIDE  
 CC CELL WALL PRECURSORS.  
 CC -1- CATALYTIC ACTIVITY: D-alanyl-D-alanine + H(2)O = 2 D-alanine.  
 CC -1- PATHWAY: Peptidoglycan synthesis; final stages.  
 CC -1- SUBCELLULAR LOCATION: Membrane-associated.  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S11. ALSO KNOWN AS THE  
 CC D-ALANYL-D-ALANINE CARBOXYPEPTIDASE 1 FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>  
 CC or send an email to [license@isb-sdb.ch](mailto:license@isb-sdb.ch)).  
 CC -----  
 DR EMBL; D26185; BA05246.1; -  
 DR EMBL; Z99104; CAB11786.1; -  
 DR EMBL; M13766; AAA22375.1; -  
 DR PIR; S66040; S66040.  
 DR MEROPS; S11.001; -  
 DR Subtilast; BG10074; dact.  
 DR InterPro; IPR001967; Ala/AlaCbpase1.  
 DR Pfam; PF00768; Peptidase\_S11; 1.

DR PRINTS: PR00725; DADACBPASE1.  
 KW Hydrolase; Carboxypeptidase; Peptidoglycan synthesis; Cell wall;  
 KW Membrane; Signal; Complete proteome.  
 FT SIGNAL 1 31  
 FT CHAIN 32 443 D-ALANYL-D-ALANINE CARBOXYPEPTIDASE.  
 FT ACT\_SITE 67 67 ACTIVATED BY PENICILLIN.  
 FT CONFLICT 100 100 E -> Q (IN REF. 3).  
 FT CONFLICT 227 227 E -> Q (IN REF. 4).  
 SQ SEQUENCE 443 AA; 48636 MW; DA6C5B0307D7C117 CRC64;  
 Query Match 10.7%; Score 90.5; DB 1; Length 443;  
 Best Local Similarity 24.1%; Pred. No. 11;  
 Matches 32; Conservative 25; Mismatches 49; Indels 27; Gaps 4;  
 QY 6 FILMDTGEVSLKPRVTVLQNGKMSSTVSEDFILPYKGLBKGYQFDGMEISG 65  
 DB 310 FEMKRIAGVGDVKGHK-TISVDKGEKVGIVTKAFSLPYKNGE-EKRYAK----- 361  
 QY 66 FEGKKDAGVIVLSKDTFIPFKFKIEKKEENKPTPVSKKDPQVHNSQLNESHK 125  
 DB 362 -----VTLKNDLTLAPYKKGTVKVK-----LTAEYTGDDQDYGFLNSDLAG 402  
 QY 126 EDLQREHHSOKSD 138  
 DB 403 VDLTKKENVKKN 415  
 RESULT 12  
 SPT7\_YEAST  
 ID SPT7\_YEAST STANDARD; PRT; 1332 AA.  
 AC P35177;  
 DT 01-FEB-1994 (Rel. 28, Created)  
 DT 01-FEB-1994 (Rel. 28, Last sequence update)  
 DT 01-OCT-1996 (Rel. 34, Last annotation update)  
 DE Transcriptional activator SPT7.  
 GN SPT7 OR YBR081C OR YBR0739.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Saccharomycetes: Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycaceae.  
 OK NCBI\_TaxId=4932;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=S288C;  
 RX MEDLINE=95229044; PubMed=7713415;  
 RA Gantheroff L.J., Dollard C., Tan P., Winston F.;  
 RT "The *Saccharomyces cerevisiae* SPT7 gene encodes a very acidic protein  
 RT important for transcription in vivo." ;  
 RL Genetics 139:523-536(1995).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=S288C;  
 RX MEDLINE=95076715; PubMed=7985423;  
 RA van der Aart O.J.M., Barthe C., Doignon F., Aigle M., Crouzet M.,  
 RA Steensma H.Y.;  
 RT "Sequence analysis of a 31 kb DNA fragment from the right arm of  
 RT *Saccharomyces cerevisiae* chromosome II." ;  
 RL Yeast 10:959-964(1994).  
 RN [3]  
 RP SEQUENCE OF 1-835 FROM N.A.  
 RC STRAIN=S288C;  
 RA Ande B., Cziepluch C., Hein C., Jauniaux J.C., Urrestarazu A.,  
 RA Visiers S.;  
 RT Submitted (Aug-1994) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP SEQUENCE OF 463-523 FROM N.A.  
 RX MEDLINE=92285152; PubMed=1350857;  
 RA Haynes S.R., Dollard C., Winston F., Beck S., Trowdale J.,  
 RA David I.B.;  
 RT "The bromodomain: a conserved sequence found in human, *Drosophila* and  
 RT yeast proteins." ;  
 RL Nucleic Acids Res. 20:2603-2603(1992).  
 CC -1- FUNCTION: TRANSCRIPTIONAL ACTIVATOR OF TY ELEMENTS AND POSSIBLY  
 CC OTHER GENES.

```

CC CC -1- SUBCELLULAR LOCATION: Nucleat.
CC CC -1- SIMILARITY: Contains 1 bromodomain.
-----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See http://www.ibt-ib.ch/announce/
CC CC or send an email to license@ibt-ib.ch).
-----
DR EMBL: L22537; AAC37424.1; -
DR EMBL: X76294; CAA53940.1; -
DR EMBL: Z35950; CAA85026.1; -
DR EMBL: M87651; AAA35087.1; -
DR PIR: S41552; S41552.
DR HSSP: Q92831; 1B91.
DR TRNSPAC: T04835; -
DR SGD: S0000285; SPT7.
DR GO: GO:0000124; C:SAGA complex; IDA.
DR InterPro: IPR001487; Bromodomain.
DR Pfam: PF00439; Bromodomain.1.
DR PRINTS: PR00503; BROMODOMAIN.
DR SMART: SM00297; BROMO; 1.
DR PROSITE: PS00633; BROMODOMAIN_1; 1.
DR PROSITE: PS50014; BROMODOMAIN_2; 1.
DR Trnascriptipon regulation: Nuclear protein; Activator; Bromodomain.
KW DOMAIN 458 528 BROMODOMAIN.
SQ SEQUENCE 1332 AA; 152616 MW; 083B63624669244F CRC64;

Query Match 10.7%; Score 90.5; DB 1; Length 1332;
Best Local Similarity 24.5%; Pred. No. 35;
Matches 39; Conservative 26; Mismatches 57; Indels 37; Gaps 8;

QY 25 VTIQNGKMSSTVSYSEDFILPVYKGLHEKTYQDPDGEHSFBECKDAGVYINISKPTFI 84
Db 549 ITIRRRADLEKEI---EDM-----BKQDYELDEBEVYAGSRKG---LNNGAMLA 594
QY 85 K---PVEFKIIEKKEEKKEKPTFD-----VSKKD-----NPGVNSQLNESH 123
Db 595 KENGKVESEKSSKTKVDAPLPNDKLTISVIPGEKEKDKTASSVTYVHEVNNKIKENG 654
QY 124 RKEDLQR-EESHOKSDSTKDVATATVLD-KNNISSEKSTTN 160
Db 655 KNEEQDWVESSKTEDESKDAKQDTEDELQDKTAEK 693

RESULT 13
LIP STAEF STANDARD; PRT; 688 AA.
ID LIP STAEF
AC Q02510.
DT 01-JUN-1994 (Rel. 29, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Lipase precursor (EC 3.1.1.3) (Glycerol ester hydrolase).
GN GELC OR S80281.
OS Staphylococcus epidermidis.
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=1282;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 303-315.
RC STRAIN=9;
RX MEDLINE=931171870; PubMed=8436947;
RA Farrell A.M., Foster T.J., Holland K.T.;
RT "Molecular analysis and expression of the lipase of Staphylococcus
epidermidis.";
RL J. Gen. Microbiol. 139:267-277 (1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 12228;
RA Zhang Y., Ren S., Li H., Fu G., Lu L., Lu G., Jia J., Tu Y., Qin Z.,
Chen Z., Wen Y.;
RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.

```

```

CC CC -1- CATALYTIC ACTIVITY: triacylglycerol + H(2)O = diacylglycerol + a
CC CC fatty acid anion.
CC CC -1- SUBCELLULAR LOCATION: Secreted.
CC CC -1- MISCELLANEOUS: THE EXPRESSION OF STAPHYLOCOCCUS LIPASE IS
CC CC NEGATIVELY REGULATED BY BACTERIOPHAGE LYSOGENIZATION (LIPASE
CC CC CONVERSION)'.
CC CC -1- SIMILARITY: BELONGS TO THE ABHYDROLASE SUPERFAMILY. LIPASE FAMILY.
CC CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See license@ibt-sib.ch).
CC CC -----
CC CC EMBL, M95577; AAA19729.1; -.
CC CC DR EMBL, AE016744; AAC03878.1; -.
CC CC DR PIR, A47705; A47705.
CC CC DR InterPro, IPR005877; Gpos_YsIRK.
CC CC DR InterPro, IPR000734; Lipase.
CC CC DR InterPro, IPR000379; Ser_estr_site.
CC CC DR Pfam, PF04650; YsIRK signal; 1.
CC CC DR TIGRfam, TIGR01168; YsIRK signal; 1.
CC CC DR PROSITE, PS00120; LIPASE_SER; 1.
CC CC KW Hydrolase; Lipid degradation; Zymogen; Signal; Complete proteome.
CC CC FT SIGNAL 1 35 POTENTIAL.
CC CC FT PROPEP 36 302 REMOVED IN THE MATURE FORM.
CC CC FT CHAIN 303 688 LIPASE.
CC CC FT ACT_SITE 418 418 CHANGE RELAY SYSTEM (BY SIMILARITY).
CC CC FT ACT_SITE 648 648 CHANGE RELAY SYSTEM (BY SIMILARITY).
CC CC FT CONFLICT 96 96 W -> L (IN REF. 1).
CC CC FT CONFLICT 120 120 E -> G (IN REF. 1).
CC CC FT CONFLICT 120 120 E -> G (IN REF. 1).
CC CC SQ SEQUENCE 688 AA; 77343 MW; 6C95DB3A78AF86F6 CRC64;

Query Match 10.6%; Score 90; DB 1; Length 688;
Best local Similarity 24.4%; Pred. No. 19;
Matches 39; Conservative 30; Mismatches 65; Indels 26; Gaps 7;

QY 9 NKDTGSEVSELPKPHAVVTYIQNGKMSSTIYSEEPFLPVYKGELEKGYQDPGWEISGFE 67
DB 68 NKNVNESKNV-----SITENESLHNETPKNEPDMW-----QQQKXSQNDNKSBSVVEQ 115
QY 68 GKXGAYVINSKDTPIKPVPKIE-ERKKEENKPTFDVSKKKONPOV--NHSQINSR 124
DB 116 NKEENAPVQNNSE--KPDQEQVELEKHSENNQTLHSAKSAQSNEDVKTTPSQLDNTAA 172
QY 125 KEDLQREHSGKSDSTQVATVYDVKNNISKSTYNNPK 164
DB 173 KQSDQKENSUKDPTQSSKTTDL-----RATNQOSK 205

RESULT 14
TIG_STAM STANDARD; PRT; 433 AA.
ID TIG_STAM STANDARD; PRT; 433 AA.
AC Q99T16;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Trigger factor (TF).
GN TIG OR SAV1.675 OR SAV1.499 OR MW1619.
OS Staphylococcus aureus (strain M50 / ATCC 700699),
OS Staphylococcus aureus (strain N315), and
OS Staphylococcus aureus (strain MW2).
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=158878, 158879, 196620;
RN
RP SEQUENCE FROM N.A.
RC STRAIN=M50 / ATCC 700699, and N315;
RA MEDLINE=21311952; PubMed=11418146;
RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,
RA Cui L., Oguchi A., Aoki K.-I., Nagai Y., Iian J.-O., Ito T.,
RA Kanamori M., Matsunaru H., Maruyama A., Murakami H., Hoshiyama A.,

```

RA Mizutani-Ui Y., Takahashi N.K., Sawano T., Inoue R.-I., Kaito C.,  
RA Sekimizu K., Hiraoka H., Kuhara S., Goto S., Yabuzaki J.,  
RA Kanehisa M., Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T.,  
RA Hattori M., Ogasawara N., Hayashi H., Hirata S.,  
RA "Whole genome sequencing of methicillin-resistant *Staphylococcus aureus*."  
RT Lancet 357:1225-1240(2001).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=MM2;  
RX MEDLINE=22040717; PubMed=12044378;  
RA Baha T., Takuchi F., Kuroda M., Yuzawa H., Aoki K.-I., Oguchi A.,  
RA Nagai Y., Iwama N., Asano K., Naimi T., Kuroda H., Cui L.,  
RA Yamamoto K., Hiramatsu K.,  
RA "Genome and virulence determinants of high virulence community-acquired MRSA."  
RT Lancet 359:1819-1827(2002).  
RL  
CC -1- FUNCTION: Involved in protein export. Acts as a chaperone by maintaining the newly synthesized protein in an open conformation (By similarity).  
CC -1- SIMILARITY: BELONGS TO THE FKBP-TYPE PPIASE FAMILY. TIG SUBFAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL: AP003363; BAB57837.1; -  
CC EMBL: AP003134; BAB42766.1; -  
CC EMBL: AP004827; BAB95484.1; -  
CC PIR: A89951; A89951.  
CC HAMAP: MF\_00303; -; 1.  
CC InterPro: IPR001179; FKBP\_PPIase.  
CC InterPro: IPR005215; Trig\_fac.  
CC Pfam: PF00254; FKBP.1.  
CC TIGRPFAM: TIGR00115; tlg.1.  
CC PROSITE: PS00453; FKBP\_PPIASE\_1; FALSE NEG.  
CC PROSITE: PS00454; FKBP\_PPIASE\_2; FALSE NEG.  
CC PROSITE: PS50059; FKBP\_PPIASE\_3; 1.  
CC K1 Cell division; Chaperone; Isomerase; Rotamase; Complete proteome.  
CC FT DOMAIN 163 248 PPIASE, FKBP-TYPE.  
CC SEQUENCE 433 AA; 48609 MW; 8865D9A6A1BC1E7 CRC64;  
  
Query Match 10.6%; Score 89.5; DB 1; Length 433;  
Best Local Similarity 22.1%; Pred. No. 12;  
Matches 50; Conservative 33; Mismatches 60; Indels 83; Gaps 13;  
  
QY 11 DTGEVSELKPHRYVT-IQNGKE--MSSTIVSEEDFLPVYKG-ELEK----- 54  
DB 81 DETDIKPAVQPEVSTQIEKQDFIEFAIVTVEPEKLGADYKGLIEIKQETELSDDELOE 140  
QY 55 -----GIQPDGWEISG--FEKCKAGYINISKQTFIKP 86  
DB 141 AIDHSIAGLAEVNVKEDGVENGDTVNIIDFSG-SVDGEHFEQGAQGVLDLEIGSGFI-P 198  
QY 87 VFK-----KIEEKE-----EE--NKPEFDVS-----KKKNPQVNHSQLNE- 121  
DB 199 GFEEQLSEKQVDEEDKDVVTPPEBYHAEELACKENAFKTKVNEIKFKEPELDTDEIANEL 258  
QY 122 -----SHRKEDLOREHESQKSDTKDVTATVLDKNINISKSTTN 160  
DB 259 DAEANTVDEYKENTKRLAEQKATDAENV-----EKEEATYATTN 299  
  
RESULT 15  
SKM1\_YEAST STRAND: PRT; 655 AA.  
AC Q12469; Q06940;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Serine/threonine-protein kinase SKM1 (EC 2.7.1.-) (Protein kinase 75490 D).  
GN SKM1 OR YOL113W OR HRA655.  
OS Saccharomyces cerevisiae (baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
OX NCBI\_TaxID=4932;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=S288C / GRF8;  
RX MEDLINE=97197189; PubMed=9044278;  
RA Martin H., Mendoza A., Rodriguez-Pachon J.M., Molina M., Nombela C.,  
RA "Characterization of SKM1, a Saccharomyces cerevisiae gene encoding a novel Ste20/Pak-like protein kinase."  
RT Mol. Microbiol. 23:431-444(1997).  
RL [2]  
RN  
RP SEQUENCE FROM N.A.  
RX MEDLINE=96076631; PubMed=7502582;  
RA Vandebol M., Durand P., Portetelle D., Hilger F.,  
RT "Sequence analysis of a 44 kb DNA fragment of yeast chromosome XV including the Ty1-H3 retrotransposon, the *sufl*(+) frameshift suppressor gene for tRNA-Gly, the yeast transfer RNA-Thr-1a and a delta element."  
RT Yeast 11:1069-1075(1995).  
RL  
CC -1- FUNCTION: MAY BE INVOLVED IN CELLULAR SIGNALING OR CYTOSKELETAL  
CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.  
CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.  
CC ST20 SUBFAMILY.  
CC -1- SIMILARITY: Contains 1 CRIB domain.  
CC -1- SIMILARITY: Contains 1 PH domain.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL: X69322; CAA49163.1; -  
CC EMBL: Z48149; CAA88147.1; -  
CC EMBL: Z74855; CAA99132.1; -  
CC PIR: S51884; S51884.  
CC SGD: S0005473; SKM1.  
CC GO: GO:0000074; P:regulation of cell cycle; IGI.  
CC InterPro: IPR000095; PakBox/RhoBinding.  
CC InterPro: IPR001849; PH.  
CC InterPro: IPR000719; Prot. Kinase.  
CC InterPro: IPR002290; Ser Thr Kinase.  
CC InterPro: IPR001245; Tyr\_kinase.  
CC Pfam: PF00786; PBD; 1.  
CC Pfam: PF00169; PH; 1.  
CC Pfam: PF00069; pkinase; 1.  
CC PRINTS: PR00109; TYRKINASE.  
CC ProDom: PD000001; Prot\_kinase; 1.  
CC SMART: SMO0285; PBD; 1.  
CC SMART: SMO0233; PH; 1.  
CC SMART: SMO0220; S\_TKC; 1.  
CC PROSITE: PS50108; CRIB; 1.  
CC PROSITE: PS50003; PH\_DOMAIN; 1.  
CC PROSITE: PS50011; PROTEIN\_KINASE\_DOM; 1.  
CC PROSITE: PS50018; PROTEIN\_KINASE\_ST; 1.  
CC K1 transferase; Serine/threonine-protein kinase; ATP-binding.  
CC FT DOMAIN 3 118 PH.  
CC DOMAIN 123 136 CRIB.  
CC FT DOMAIN 360 639 PROTEIN KINASE.  
CC FT NP\_BIND 366 374 ATP (BY SIMILARITY).  
CC FT BINDING 406 406 ATP (BY SIMILARITY).  
CC FT ACT\_SITE 507 507 BY SIMILARITY.  
CC FT CONFLICT 303 303 S -> A (IN REF. 1).  
CC FT CONFLICT 306 306 R -> K (IN REF. 1).  
CC FT CONFLICT 320 320 H -> Y (IN REF. 1).

```

RESULT 16
IGA2_HAFLIN
ID_IGA2_HAFLIN STANDARD: PRT; 1702 AA.
AC P45384;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 26-FEB-2003 (Rel. 41, Last annotation update)
DE Immunoglobulin A1 protease precursor (EC 3.4.21.72) (IGA1 protease) .
GN IGA1.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HK715 / Serotype B;
RX MEDLINE=92234949; Pubmed=1373717;
RA Poulson K., Reinholdt J., Kilian M.;
RT "A comparative genetic study of serologically distinct Haemophilus
influenzae type 1 immunoglobulin A1 proteases.",
RJ J. Bacteriol. 174:2913-2921(1992).
CC -1- FUNCTION: VIRULENCE FACTOR; CLEAVES HOST IMMUNOGLOBULIN A
CC -1- PRODUCING INTRACT PC AND FAB FRAGMENTS.
CC -1- CATALYTIC ACTIVITY: Cleavage of immunoglobulin A molecules at
CC certain Pro-|-Xaa bonds in the hinge region. No small molecule
CC substrates are known.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- DOMAIN: THE SIGNAL PEPTIDE GUIDES THE PRECURSOR TO THE PERIPLASMIC
CC SPACE. AND THE CARBOXY-TERMINAL HELPER DOMAIN ASSOCIATES WITH THE
CC OUTER MEMBRANE TO FORM A PORE FOR EXPORTATION BY THE PROTEASE
CC DOMAIN. THE HELPER DOMAIN IS THEN RELEASED BY AUTOPROTEOLYSIS (BY
CC SIMILARITY).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S6.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on ways
CC use by non-profit institutions as long as their content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/ib-sib.ch).
CC or send an email to license@ib-sib.ch).
CC -----
CC EMBL; M87489; AAA24966.1; -.
CC PIR; A41859; A41859.
CC MEROPS; S06.001; -.
CC InterPro; IPR006315; Autotransport.
CC InterPro; IPR005546; Autotransporter.
CC InterPro; IPR000710; IGA_S6.
CC InterPro; IPR004899; Pertactin.
CC Pfam; PF03797; Autotransporter; 1.
CC Pfam; PF02395; IGA1; 1.

```

	RESULT 17			
YF8	YEAST			
ID	YF8	YEAST	STANDARD;	PRT; 778 AA.
AC	P43610;			
DT	01-NOV-1995	(Rel. 32, Created)		
PT	01-NOV-1995	(Rel. 32, Last sequence update)		
DE	26-FEB-2003	(Rel. 41, Last annotation update)		
DE	Hypothetical 88.7 kDa helicase in CDC26-SAP15 intergenic region.			
GN	yfr038w.			
OS	Saccharomyces cerevisiae (Baker's yeast).			
OC	Bukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;			
OC	Saccharomycetales; Saccharomycetaceae; Saccharomyces.			
OX	NCBI_TaxID=4932;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=S288C / AB972;			
EX	MEDLINE=95400292; PubMed=7670463;"			
RA	Murakami Y., Natou M., Hagiwara H., Shibata T., Ozawa M.,			
RA	Saanuma S.-I., Saanuma M., Tsuchiya Y., Soeda E., Yokoyama K.,			
RA	Yamazaki M., Tashiro H., Eki T.;"			
RT	"Analysis of the nucleotide sequence of chromosome VI from			
RT	Saccharomyces cerevisiae.";			
RL	Nat. Genet. 10:261-268(1995).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=S288C / AB972;			
EX	MEDLINE=96287654; PubMed=8686381;			
RA	Eki T., Natou M., Hagiwara H., Abe M., Ozawa M., Saanuma S.-I.,			
RA	Tsuchiya Y., Shibata T., Watanabe K., Ono A.,			
RA	Yamazaki M.-A., Tashiro H., Hanoka F., Murakami Y.;"			
RT	"Fifteen open reading frames in a 30.8 kb region of the right arm of			
RT	chromosome VI from Saccharomyces cerevisiae."			
YL	Yeast 12:177-190(1996).			
CC	-1- SUBCELLULAR LOCATION: Nuclear (Potential).			
CC	-1- SIMILARITY: BELONGS TO THE SNF2/RAD54 HELICASE FAMILY.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See <a href="http://www.isb.ch/announce/">http://www.isb.ch/announce/</a>			
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).			
CC	-----			
DR	EMBL; D50617; AAA09277.1; --			
DR	PIR; S56293; S56293.			

DR SGD; S0001934; YPR038W.  
 DR InterPro; IPR001410; DEAD.  
 DR InterPro; IPR001650; Helicase\_C.  
 DR InterPro; IPR000330; SNF2\_N.  
 DR Pfam; PF00271; Helicase\_C; 1.  
 DR Pfam; PF00176; SNF2\_N; 1.  
 DR SMART; SM00487; DEXDC; 1.  
 DR SMART; SM00490; HELIC\_C; 1.  
 DR Hypothetical protein; Nuclear protein; DNA-binding; Helicase;  
 KW ATP-binding.  
 KM NP BIND  
 FT SITE 247 254 ATP (POTENTIAL).  
 FT SITE 352 355 DEGH BOX.  
 SQ SEQUENCE 778 AA; 88730 MW; 3BEC0857B5EAD84 CRC64;

Query Match 10.3%; Score 87; DB 1; Length 778;  
 Best Local Similarity 24.2%; Pred. No. 35;  
 Matches 44; Conservative 26; Mismatches 70; Indels 42; Gaps 6;

QY 4 KEFLINKTGEVSLKPHRYVTITONGKEMST-----IVSEDFLLPYKGELENG 55  
 DB 35 REKEVNDITADISDSDDSDSDKNGKGDNDTAPITWLDVHSDEDI----- 82  
 QY 56 YQPCWEISGFEKGDAGVIVLTKDTFKVYFKKIEKKSEENKPTFVSKKKDNPQVN 115  
 DB 83 -QLDESDSDTEAVQ-AQVVDKLANDT--KSEQKSIDDELSEMDTKVSLKIKLKNIEFVR 138  
 QY 116 HSQLES-----HRKEDLOR-----EHSQKSDSTKVATATVADKNNISSKS 157  
 DB 139 QSQVSSIIADTLTLHRSNEMVANANTKDNSDDEHSSKRRKTKKSLIDPFKQKQED 198  
 QY 158 TT 159  
 DB 199 TT 200

## RESULT 18

ID UN89\_CABEL STANDARD; PRT; 6632 AA.

AC 001761; Q17362;  
 DT 15-SEP-2003 (Rel. 42, Created)  
 DT 15-SEP-2003 (Rel. 42, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Muscle M-line assembly protein unc-89 (Uncoordinated protein 89).  
 GN UNC-89 OR C09D1.1.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;  
 OC Rhabditidae; Peloderinae; Caenorhabditis.  
 OX NCBI\_TaxID=6239;  
 RN RN  
 RP SEQUENCE FROM N.A., FUNCTION, AND TISSUE SPECIFICITY.  
 RC STRAIN-Bristol N2;  
 RX MEDLINE=96180278; PubMed=8603916;  
 RA Benian G.M., Tinley T.L., Tang X., Borodovsky M.;  
 RT "The Caenorhabditis elegans gene unc-89, required for muscle M-line  
 assembly, encodes a giant modular protein composed of Ig and signal  
 transduction domains";  
 RT J. Cell Biol. 132:835-848(1996).  
 RL (2)  
 RN RN  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-Bristol N2;  
 RA Du Z., Le T.T., Wilson R.;  
 RT Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.  
 RL (3)  
 RN RN  
 RP REVISIONS.  
 RA Waterston R.;  
 RL Submitted (Apr-2002) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: Structural component of the muscle M-line. Myofilament  
 lattice assembly begins with positional cues laid down in the  
 basement membrane and muscle cell membrane. UNC-89 responds to  
 these signals, localizes, and then participates in assembling an  
 M-line.  
 CC -1- TISSUE SPECIFICITY: Localizes to the middle of A-bands.  
 CC -1- SIMILARITY: Contains 1 DBL-homology (DH) domain.

CC -1- SIMILARITY: Contains 1 fibronectin type III domain.  
 CC -1- SIMILARITY: Contains 49 immunoglobulin-like C2-type domains.  
 CC -1- SIMILARITY: Contains 1 PH domain.  
 CC -1- SIMILARITY: Contains 5 RCD domains.  
 CC -1- SIMILARITY: Contains 1 SH3 domain.  
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC EMBL; U33058; AAB0542.1; -.  
 CC EMBL; AF003131; AAB54132.2; -.  
 CC PDB; 1FHO; 20-DEC-00.  
 DR WormBep; C09D1.1; CE30426.  
 DR InterPro; IPR003961; FN\_III.  
 DR InterPro; IPR007110; Ig-like.  
 DR InterPro; IPR003598; Ig\_C2.  
 DR InterPro; IPR003006; Ig\_MHC.  
 DR InterPro; IPR001849; PH.  
 DR InterPro; IPR007850; RCD.  
 DR InterPro; IPR000219; RhGEF.  
 DR InterPro; IPR01452; SH3.  
 DR Pfam; PF00041; fn3; 1.  
 DR Pfam; PF00047; Ig\_47.  
 DR Pfam; PF00169; PH; 1.  
 DR Pfam; PF05177; RCD; 5.  
 DR Pfam; PF00621; RhGEF; 1.  
 DR Pfam; PF00018; SH3; 1.  
 DR SMART; SM00408; IGC2; 23.  
 DR SMART; SM00325; RhGEF; 1.  
 DR SMART; SM00325; SH3; 1.  
 DR PROSITE; PS50010; DH\_2; 1.  
 DR PROSITE; PS50003; PH\_DOMAIN; 49.  
 DR PROSITE; PS50002; SH3; 1.  
 KW Muscle protein; Immunoglobulin domain; Repeat; SH3 domain;  
 KW 3D-structure.  
 FT 63 127 SH3.  
 FT DOMAIN 152 330 DH.  
 FT DOMAIN 342 498 PH.  
 FT DOMAIN 547 633 IG-LIKE C2-TYPE 1.  
 FT DOMAIN 648 736 IG-LIKE C2-TYPE 2.  
 FT DOMAIN 748 838 IG-LIKE C2-TYPE 3.  
 FT DOMAIN 946 1033 IG-LIKE C2-TYPE 4.  
 FT DOMAIN 1044 1132 IG-LIKE C2-TYPE 5.  
 FT DOMAIN 1140 1227 IG-LIKE C2-TYPE 6.  
 FT DOMAIN 1272 1315 THR-RICH.  
 FT DOMAIN 1375 1475 RCD 1.  
 FT DOMAIN 1479 1585 RCD 2.  
 FT DOMAIN 1597 1695 RCD 3.  
 FT DOMAIN 1700 1799 RCD 4.  
 FT DOMAIN 1800 1860 RCD 5.  
 FT DOMAIN 1862 2067 IG-LIKE C2-TYPE 7.  
 FT DOMAIN 2071 2163 IG-LIKE C2-TYPE 8.  
 FT DOMAIN 2171 2261 IG-LIKE C2-TYPE 9.  
 FT DOMAIN 2269 2359 IG-LIKE C2-TYPE 10.  
 FT DOMAIN 2367 2455 IG-LIKE C2-TYPE 11.  
 FT DOMAIN 2463 2564 IG-LIKE C2-TYPE 12.  
 FT DOMAIN 2563 2651 IG-LIKE C2-TYPE 13.  
 FT DOMAIN 2657 2746 IG-LIKE C2-TYPE 14.  
 FT DOMAIN 2754 2858 IG-LIKE C2-TYPE 15.  
 FT DOMAIN 2867 2980 IG-LIKE C2-TYPE 16.  
 FT DOMAIN 2994 3081 IG-LIKE C2-TYPE 17.  
 FT DOMAIN 3087 3183 IG-LIKE C2-TYPE 18.  
 FT DOMAIN 3189 3280 IG-LIKE C2-TYPE 19.  
 FT DOMAIN 3286 3376 IG-LIKE C2-TYPE 20.  
 FT DOMAIN 3384 3469 IG-LIKE C2-TYPE 21.  
 FT DOMAIN 3482 3572 IG-LIKE C2-TYPE 22.  
 FT DOMAIN 3580 3667 IG-LIKE C2-TYPE 23.

```

FT DOMAIN 3686 3777 IG-LIKE C2-TYPE 24.
PT DOMAIN 3817 3908 IG-LIKE C2-TYPE 25.
PT DOMAIN 3920 4009 IG-LIKE C2-TYPE 26.
PT DOMAIN 4018 4106 IG-LIKE C2-TYPE 27.
PT DOMAIN 4201 4297 IG-LIKE C2-TYPE 28.
PT DOMAIN 4302 4387 IG-LIKE C2-TYPE 29.
PT DOMAIN 4400 4485 IG-LIKE C2-TYPE 30.
PT DOMAIN 4489 4580 IG-LIKE C2-TYPE 31.
PT DOMAIN 4681 4771 IG-LIKE C2-TYPE 32.
PT DOMAIN 4873 4961 IG-LIKE C2-TYPE 33.
PT DOMAIN 4965 5057 IG-LIKE C2-TYPE 34.
PT DOMAIN 5067 5160 IG-LIKE C2-TYPE 35.
PT DOMAIN 5171 5260 IG-LIKE C2-TYPE 36.
PT DOMAIN 5277 5366 IG-LIKE C2-TYPE 37.
PT DOMAIN 5383 5472 IG-LIKE C2-TYPE 38.
PT DOMAIN 5487 5578 IG-LIKE C2-TYPE 39.
PT DOMAIN 5595 5685 IG-LIKE C2-TYPE 40.
PT DOMAIN 5701 5790 IG-LIKE C2-TYPE 41.
PT DOMAIN 5815 5904 IG-LIKE C2-TYPE 42.
PT DOMAIN 5925 6014 IG-LIKE C2-TYPE 43.
PT DOMAIN 6038 6130 IG-LIKE C2-TYPE 44.
PT DOMAIN 6150 6239 IG-LIKE C2-TYPE 45.
PT DOMAIN 6275 6368 IG-LIKE C2-TYPE 46.
PT DOMAIN 6413 6502 IG-LIKE C2-TYPE 47.
PT DOMAIN 6507 6596 IG-LIKE C2-TYPE 48.
PT DOMAIN 6621 6710 IG-LIKE C2-TYPE 49.
PT DISULFID 2908 2975 POTENTIAL.
PT DISULFID 3015 3065 POTENTIAL.
PT DISULFID 3707 3759 POTENTIAL.
PT DISULFID 3826 3890 POTENTIAL.
PT DISULFID 5092 5157 POTENTIAL.
PT DISULFID 5298 5350 POTENTIAL.
PT DISULFID 5508 5560 POTENTIAL.
PT DISULFID 5616 5669 POTENTIAL.
PT DISULFID 5722 5764 POTENTIAL.
PT DISULFID 5836 5901 POTENTIAL.
PT DISULFID 5946 6011 POTENTIAL.
PT DISULFID 6036 6171 POTENTIAL.
PT DISULFID 6421 6486 POTENTIAL.
PT CONFLICT 2137 2137 A -> P (IN REF. 1).
PT CONFLICT 2245 2247 A -> P (IN REF. 1).
PT CONFLICT 2258 2258 E -> G (IN REF. 1).
PT CONFLICT 2284 2284 M -> I (IN REF. 1).
PT CONFLICT 2297 2297 A -> G (IN REF. 1).
PT CONFLICT 3351 3351 DAGEY -> RRRRI (IN REF. 1).
PT CONFLICT 3884 3884 A -> V (IN REF. 1).
PT CONFLICT 3929 3929 A -> P (IN REF. 1).
PT CONFLICT 5134 5134 T -> S (IN REF. 1).
PT CONFLICT 5145 5145 G -> A (IN REF. 1).
PT CONFLICT 5185 5185 K -> N (IN REF. 1).
PT CONFLICT 5199 5199 L -> F (IN REF. 1).
PT CONFLICT 5202 5202 F -> L (IN REF. 1).
PT CONFLICT 5213 5213 A -> G (IN REF. 1).
PT CONFLICT 6178 6178 K -> E (IN REF. 1).
PT CONFLICT 6268 6268 K -> E (IN REF. 1).
SQ SEQUENCE 6632 AA; 731665 MW; 2620DIED62960E89 CRC64;

```

Query Match 10.3%; Score 87; DB 1; Length 6632;  
 Best Local Similarity 21.8%; Pred. No. 3.5e+02;  
 Matches 37; Conservative 33; Mismatches 80; Indels 20; Gaps 7;

```

QY 1 TTVKFFILKQGVSELEKPRVTATVIONGKEMSSITVEEDP--ILPYKGELEKGYQF 58
DB 1127 TKAKLTVOCKPAPRPFDRP--VSLTVEKGE--AVFAHAHFGIPLPTTMSVNGRKR 1181
QY 59 DQME--ISGFEGKQDAGVIMLSKDTFKIPV----FKXIEKK--EEENKPTFVSKKK 109
DB 1182 DQEGARVTRDESTVDASILITDTRATYSEVNHLLISVAVNTLGAERTGAQLTIEPRK 1241
QY 110 DNPQVNHSGQNSHREKEDLOREHSHQKSDSTQVATATVLDKNNISKSTT 159

```

```

DB 1242 BSVVEKQDLSSS----BVQKEIAQVKEASPEATTITMETSITSTKTT 1287

RESULT 19
ID NSB1_MOUSE STANDARD; PRT; 406 AA.
AC Q9UJ35; O88832; Q8VC71; Q9CUM1;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Nucleosome binding protein 1 (Nucleosome binding protein 45) (NBP-45) (GARP45 protein).
GN NSBP1 OR GARP45.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NCBI_TaxID=10090;

[1]
RP SEQUENCE FROM N.A., FUNCTION, SUBCELLULAR LOCATION, TISSUE SPECIFICITY, AND DEVELOPMENTAL STAGE.
RX MEDLINE=20158948; PubMed=10692437;
RA Shirakawa H., Landsman D., Postnikov Y.V., Bustin M.;
RT "NBP-45, a novel nucleosomal binding protein with a tissue-specific and developmentally regulated expression."
RJ J. Biol. Chem. 275:6368-6374(2000).

[2]
RP SEQUENCE FROM N.A.
RA Onoda G., Suzuki N., Saito H., Honda T., Sato H., Kuwano R.;
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.

[3]
RP SEQUENCE FROM N.A.
RC TISSUE=Colon;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G., Klausner R.D., Collins P.S., Wagner L., Shenmen C.M., Schuler G.D., Altschul S.F., Zeeberg B., Buettow K.H., Schaefer C.F., Bhat N.K., Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F., Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L., Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Schetz T.E., Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C., Raha S.S., Loughellano N.A., Peters G.J., Abramson R.D., Mullaly S.J., Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H., Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hultk S.W., Vallalath D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A., Fahy J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A., Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G., Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C., Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butlerfield Y.S.N., Krzywinski M.I., Skalska U., Smalhus D.E., Scherch A., Schein J.B., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences."
RJ Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).

[4]
RP SEQUENCE OF 7-195 FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Hippocampus;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y., Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamada S., Saito T., Okazaki Y., Gojobori T., Bono H., Kaizawa T., Saito R., Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T., Pleischmann W., Gaasterland T., Glass C., King B., Kochiwa H., Kuehl P., Lewis S., Matsumoto Y., Nikaide I., Pesole G., Quackenbush J., Schmiel L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Mashio T., Sakai K., Okido T., Furuno M., Aono H., Balarelli R., Barsch G., Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldi M.F., Brownstein M.J., But C., Fletcher C., Fujita M., Gariboldi M., Guerninich S., Hill D., Hofmann M., Hume D.A., Kamuya M., Lee N.H., Lyons P., Marchionni L., Meshima J., Mazaralli J., Mombarts P., Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N., Sasaki H., Sato K., Schoenbach C., Seta T., Shibata Y., Storch K.-F., Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,

```

RA Wyszaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohseuki S.,  
 RA Hayashizaki Y.:  
 RT "Functional annotation of a full-length mouse cDNA collection.";  
 RL Nature 409:685-690(2001).  
 CC -1- FUNCTION: Binds specifically to nuclear core particles and acts as  
 CC a transcriptional activator.  
 CC -1- SUBCELLULAR LOCATION: Nuclear.  
 CC -1- TISSUE SPECIFICITY: Widely expressed with highest levels in  
 CC submaxillary gland, thymus, kidney and liver and lowest levels in  
 CC brain, lung, pancreas and eye.  
 CC -1- DEVELOPMENTAL STAGE: Highest levels are found in 7-day-old  
 CC embryos. Levels in the 7-day-old embryo are 4-fold higher than in  
 CC the adult and almost 10-fold higher than in later embryonic  
 CC stages.  
 CC -1- SIMILARITY: BELONGS TO THE HMG-14/HMG-17 PROTEIN FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: AF23454; AAF30179.1; -;  
 CC EMBL: AB018374; BAA33783.2; -;  
 CC EMBL: BC021626; AAB21626.1; -;  
 CC EMBL: AK013748; BAB28982.1; -;  
 CC MGD: MGI:1355295; Nsdpl.  
 CC DR GO: 0005654; C:nucleoplasm; IDA.  
 CC DR GO: 0003682; P:chromatin binding activity; IDA.  
 CC DR GO: 0006356; P:regulation of transcription from Pol I prom. .; IDA.  
 CC DR InterPro: IPR000079; HMG\_14\_17.  
 CC DR Pfam: PF01101; HMG14\_17; 1.  
 CC DR SMART: SM00527; HMG17; 1.  
 CC DR PROSITE: PS00355; HMG14\_17; FALSE NEG.  
 CC KM Transcription regulation; Activator; DNA-binding; Nuclear protein.  
 CC FT CONFLICT 74 M -> V (IN REF. 1).  
 CC FT CONFLICT 390 N -> H (IN REF. 2).  
 CC SQ SEQUENCE 406 AA; 45344 MW; 59AA305613BC9679 CRC64;  
 CC  
 CC Query Match 10.2%; Score 86.5; DB 1; Length 406;  
 CC Best Local Similarity 28.0%; Pred. No. 19;  
 CC Matches 44; Conservative 16; Mismatches 52; Indels 45; Gaps 8;  
 CC  
 CC QY 9 NMDTGEVSLKRRVTYVTTQNGK-----EHSSTTVSEEDFLPYKGLKGYPDGM 61  
 CC DB 141 HNDTGE-----EVDGKIEEGGLNEKPKGTAKSED---AEYSKDEEEKDNEKG- 185  
 CC  
 CC QY 62 EISGEGKKDAGVNLKSD-----TFIKPVKKIIEK---KEENKFTPVYSKK 108  
 CC DB 186 -----EDGKEEDDEKEEKEDDGGDTGTEKEVEQNKKEARBDDGCKCKEENK---EVGKE 237  
 CC  
 CC QY 109 KDNPNVNSQLNESHKEDLQREE--HSQKSDSTKDV 143  
 CC DB 238 GQPEEDGKEDLHREVGKEDLHREDDGKGGPEEDGKEI 274  
 CC  
 CC RESULT 20  
 CC RAT1\_YEAST STANDARD; PRT; 1006 AA.  
 CC ID RAT1\_YEAST  
 CC AC Q02792;  
 CC DT 01-OCT-1993 (Rel. 27, Created)  
 CC DT 01-OCT-1993 (Rel. 27, Last sequence update)  
 CC DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 CC DB Ribonucleic acid trafficking protein 1 (5'-3' exoribonuclease)  
 CC DB (EC 3.1.11.-) (P116).  
 CC GN RAT1 OR HKE1 OR TAP1 OR YOR048C.  
 CC OS Saccharomyces cerevisiae (Baker's yeast).  
 CC OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 CC OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.  
 CC OX NCBI\_TaxID=4932;  
 CC RN (1)

RP SEQUENCE FROM N.A.  
 RX MEDLINE=92331925; PubMed=1628825;  
 RA Amberg D.C., Goldstein A.L., Cole C.N.;  
 RT "Isolation and characterization of RAT1: an essential gene of  
 RT Saccharomyces cerevisiae required for the efficient nucleocytoplasmic  
 RT trafficking of mRNA."  
 RL Genes Dev. 6:1173-1189(1992).  
 RN [2]  
 RN SEQUENCE FROM N.A., AND EXORIBONUCLEASE ACTIVITY.  
 RX MEDLINE=93109318; PubMed=8417335;  
 RA Kenna M., Stevens A., McCammon M., Douglas M.G.;  
 RT "An essential yeast gene with homology to the exonuclease-encoding  
 RT XEN1/KEN1 gene also encodes a protein with exoribonuclease  
 RT activity."  
 RL Mol. Cell. Biol. 13:341-350(1993).  
 RN [3]  
 RN SEQUENCE FROM N.A.  
 RX MEDLINE=93268292; PubMed=8497260;  
 RA Aldrich T.L., di Segni G., McConaughy B.L., Xeen N.J., Whelen S.,  
 RA Hall B.D.;  
 RP "Structure of the yeast TAP1 protein: dependence of transcription  
 RT activation on the DNA context of the target gene."  
 RL Mol. Cell. Biol. 13:3434-3444(1993).  
 RN [4]  
 RN SEQUENCE FROM N.A.  
 RA Landt O., Hiesel R., Unseld M.;  
 RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RA Bohn C., Bolotin-Pukhara M., Dalgman-Fornier B., Dang D.V.,  
 RA Valens M.;  
 RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: May function in the processing and/or trafficking of  
 CC nuclear mRNA. May be involved in general transcription as well.  
 CC Possesses 5'-3' exoribonuclease activity degrading poly(A) to  
 CC mainly 5'-AMP.  
 CC -1- CORFACTOR: Requires magnesium.  
 CC -1- SUBCELLULAR LOCATION: Nuclear.  
 CC -1- SIMILARITY: Belongs to the 5'-3' exonuclease family.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: S61567; AAB26818.1; -;  
 CC EMBL: M95626; AAA34960.1; -;  
 CC EMBL: L06011; AAA16950.1; -;  
 CC EMBL: Z11746; -; NOT ANNOTATED\_CDS.  
 CC EMBL: Z74956; CAA9240.1; -;  
 CC PIR: S20126; S20126.  
 CC DR SGD: S0003574; RAT1.  
 CC DR GO: 0005634; C:nucleus; IDA.  
 CC DR GO: 0004534; P:5'-3' exoribonuclease activity; IDA.  
 CC DR GO: 0006365; P:35S primary transcript processing; IMP.  
 CC DR GO: 0006396; P:RNA processing; IMP.  
 CC DR InterPro: IPR004859; Ptc\_35exo.  
 CC DR Pfam: PF03159; XRN\_N; 1.  
 CC KM Nuclear protein; Hydrolyase; Nuclease; Exonuclease; Repeat.  
 CC FT DOMAIN 525 528 POLY-GLU.  
 CC FT DOMAIN 793 797 POLY-ASN.  
 CC FT DOMAIN 955 999 CONTAINS 2 X SRVD, 2 X NNTY, AND 2 X  
 CC YSGN REPEATS.  
 CC FT MUTAGEN 683 683 Y->H: IN ALLELE TAP1-1; ACTIVATES  
 CC TRANSCRIPTION OF THE PROMOTER-DEFECTIVE  
 CC YEAST SUP4 TRNA(TYR) ALLELE SUP4AS3161.  
 CC FT SEQUENCE 1006 AA; 115933 MW; 5DDDB80245F3E12A CRC64;  
 CC SQ  
 CC Query Match 10.2%; Score 86.5; DB 1; Length 1006;  
 CC Best Local Similarity 19.0%; Pred. No. 50;





"The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.";  
 Nature 390:249-256 (1997).  
 [4]  
 GROWTH REQUIREMENTS.  
 STRAIN-168;  
 MEDLINE=98389671; PubMed=9721295;  
 Murray T., Popham D.L., Setlow P.;  
 "Bacillus subtilis cells lacking penicillin-binding protein 1 require increased levels of divalent cations for growth."  
 J. Bacteriol. 180:4555-4563 (1998).  
 [5]  
 SUBCELLULAR LOCATION.  
 STRAIN-168;  
 MEDLINE=99255546; PubMed=10322023;  
 Pedersen L.B., Angert E.R., Setlow P.;  
 "Septal localization of penicillin-binding protein 1 in *Bacillus subtilis*."  
 J. Bacteriol. 181:3201-3211 (1999).  
 -1- FUNCTION: CELL WALL FORMATION. SYNTHESIS OF CROSS-LINKED PEPTIDOGLYCAN FROM THE LIPID INTERMEDIATES. THE ENZYME HAS A PERICILLIN-INSENSITIVE TRANSGLYCOSYLASE N-TERMINAL DOMAIN (FORMATION OF LINEAR GLYCAN STRANDS) AND A PERICILLIN-SENSITIVE TRANSEPTIDASE C-TERMINAL DOMAIN (CROSS-LINKING OF THE PEPTIDE SUBUNITS) (BY SIMILARITY).  
 -1- PATHWAY: Peptidoglycan synthesis, final stages.  
 -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. PROBABLY FOUND ALL OVER THE WHOLE CELL AT LOW CONCENTRATIONS. ALSO LOCALIZES TO THE DIVISION SITE IN VEGETATIVE CELLS.  
 -1- DEVELOPMENTAL STAGE: EXPRESSION IS CONSTANT DURING GROWTH, DECREASES DURING SPOGULATION AND IS INDUCED APPROXIMATELY 15 MIN INTO SPORE GERMINATION.  
 -1- PTM: THE PRODUCT EXPRESSED FROM THE TRANSLATION OF THE PONA GENE APPEARS AS TWO BANDS ON A GEL (1A AND 1B), BUT THE SPECIFIC AMINO ACID SEQUENCE OF EACH PROTEIN IS UNKNOWN.  
 -1- PTM: THE N-TERMINUS IS BLOCKED.  
 -1- MISCELLANEOUS: CELLS LACKING THE PROTEIN REQUIRE INCREASED LEVELS OF MG(2+) OR CA(2+) FOR GROWTH AND GERMINATION. APPROXIMATELY 50% OF CELLS WITHOUT THE PROTEIN CONTAIN ABNORMAL FTSG RINGS, SUGGESTING IT IS INVOLVED IN SEPTUM SYNTHESIS. INCREASED LEVELS OF MG(2+) OR CA(2+) ONLY PARTIALLY ELIMINATE THE SEPTATION DEFECTS.  
 -1- SIMILARITY: Contains 1 fibronectin type III domain.  
 -1- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE TRANSGLYCOSYLASE FAMILY.  
 -1- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO THE TRANSEPTIDASE FAMILY.  
 -----  
 This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 -----  
 EMBL; U11883; AAA64947.1; -;  
 EMBL; L47838; AAB38459.1; -;  
 EMBL; Z99115; CAB14148.1; -;  
 PIR; I40529; I40529.  
 Subtilist; BG10954; PONA.  
 InterPro; IPR003961; FN III.  
 InterPro; IPR001264; Glyco trans 51.  
 InterPro; IPR001460; Transpeptidase.  
 Pfam; PF00041; fn3; 1.  
 Pfam; PF00912; Transglycosyl; 1.  
 Pfam; PF00905; Transpeptidase; 1.  
 ProDom; PD001895; Glyco\_trans\_51; 1.  
 SMART; SM00060; FN3; 1.  
 Peptidoglycan synthesis; Cell wall; Transferase; Glycosyltransferase;  
 Hydrolyase; Multifunctional enzyme; Transmembrane; Signal-anchor;  
 Antibiotic resistance; Complete proteome.  
 CYTOPLASMIC (POTENTIAL).  
 TRANSMEM 38 58 SIGNAL-ANCHOR (TYPE II MEMBRANE PROTEIN)

FT DOMAIN 59 914 (POTENTIAL).  
 FT DOMAIN 77 246 EXTRACELLULAR (POTENTIAL).  
 FT DOMAIN 329 662 TRANSGLYCOSYLASE.  
 FT DOMAIN 706 784 TRANSEPTIDASE.  
 FT ACT SITE 390 390 FIBRONECTIN TYPE-III.  
 SQ SEQUENCE 914 AA; 99562 MW; 6978B33DFE2423B6 CRC64;  
 Query Match 10.1%; Score 86; DB 1; Length 914;  
 Best Local Similarity 23.0%; Pred. No. 49;  
 Matches 45; Conservative 33; Mismatches 56; Indels 62; Gaps 9;  
 QY 3 VKETFLNKDTGEVSELPKPRVTVTITQNGKMSSTVSEEDFLPYKKELEKGYQPDGME 62  
 DB 737 VKQSV---DGGSYSR-----IQNS-----AKAAVIGVQPGSV---YKFEVTA 774  
 QY 63 ISGFGKDA--GYINLSKDTFTKPKIKKEKEENK----- 100  
 DB 775 VSD-DGKSTASTSYVPPAARDEDEKDDQQTDEKQDDTQTGDDSQKDDGQTGD 833  
 QY 101 ----PTFPVSKKKNPQVNSQLN-----ESHKEDLQREHSQKSDSTKDVTA 145  
 DB 834 QTDSTNDQDKKQDDTNTNPSDNNQDOSNDNDNNSNNQDTSQDSNGKDDSTG--SD 891  
 QY 146 TVLDKNTSSKSTNN 161  
 DB 892 TKNKNTTSNKTQNS 907  
 RESULT 23  
 MDN1\_HUMAN STANDARD; PRT; 5596 AA.  
 AC Q9N022; O15019;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Midasin (MIDAS-containing protein).  
 GN MDN1 OR KIAA0301.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Butheraia; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Testis;  
 RX PubMed=12102729;  
 RA Garbaxino J.E., Gibbons I.R.;  
 RT "Expression and genomic analysis of midasin, a novel and highly conserved AAA protein distantly related to dynein."  
 RL BMC Genomics 3:18-18(2002).  
 RN [2]  
 RP SEQUENCE OF 1255-2356 AND 3550-5596 FROM N.A.  
 RC TISSUE=Testis;  
 RA Tracey A.;  
 RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE OF 3550-5596 FROM N.A.  
 RC TISSUE=Brain;  
 RX MEDLINE=97349984; PubMed=9205841;  
 RA Nagase T., Ishikawa K.-I., Nakajima D., Ohira M., Seki N., Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;  
 RT "Prediction of the coding sequences of unidentified human genes. VII. The complete sequences of 100 new cDNA clones from brain which can code for large proteins in vitro."  
 RL DNA Res. 4:141-150(1997).  
 CC -1- FUNCTION: May function as a nuclear chaperone and be involved in the assembly/disassembly of macromolecular complexes in the nucleus.  
 CC -1- SUBCELLULAR LOCATION: Nuclear (By similarity).  
 CC -1- SIMILARITY: Contains 1 WFA domain.  
 CC -----  
 This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation -

```

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL AF503992; AAM77722.1; -
DR EMBL AL096678; CAB86660.1; -
DR EMBL AL096678; CAB86660.1; -
DR EMBL AB002299; BAA20761.1; -
DR Genew; HGNC:18302; MDN1.
DR GO; GO:0005634; C:nucleus; NAS.
DR GO; GO:0016887; F:ATPase activity; NAS.
DR GO; GO:0003754; F:chaperone activity; NAS.
DR GO; GO:0006461; P:protein complex assembly; NAS.
DR InterPro; IPR003593; AAA_ATPase.
DR InterPro; IPR002035; VWF_A.
DR Pfam; PF00004; AAA; 1.
DR SMART; SM00382; AAA; 7.
DR SMART; SM00327; VWA; 1.
DR PROSITE; PS50234; VWFA; 1.
DR Chaperone; ATP-binding; Repeat; Nuclear protein.
KM NP_BIND 329 336
FT NP_BIND 677 684
FT NP_BIND 1084 1091
FT NP_BIND 1390 1397
FT NP_BIND 1753 1760
FT NP_BIND 2066 2073
FT NP_BIND 3566 3573
FT DOMAIN 4784 4791
FT DOMAIN 5008 5013
FT DOMAIN 5182 5187
FT DOMAIN 5384 5387
FT CONFLICT 2287 2312
SQ SEQUENCE 5596 AA; 632802 MW; 586C62616A1F96D4 CRC64;
Query Match
Best Local Similarity 23.4%; Score 86; DB 1; Length 5596;
Matches 30; Conservative 27; Mismatches 63; Indels 8; Gaps 2;
QY 18 LKPARVTYVTONGKEMSTIVSEBDFILPVYKGELEKG-----YQDGMETSGEFGKQD 71
DB 4629 LATHSTAKLSVLAQVFTELAQKQFCFLKEMEDSAGAGATBFHYEGGIGEGGKMD 4688
QY 72 AGVNLISKDPIKPVFKKIEKEKEENKPTFVSKKKNPVNHSQNLNESHKREKLDRE 131
DB 4689 VS--DQIGBEQVETPFQKQKEDPDPSKSDIKGDNALNEMSEDPFQKMDGELBQD 4746
QY 132 EHSQKSDS 139
DB 4747 EDDKSDS 4754
RESULT 24
DRS1_YEAST STANDARD; PRT; 752 AA.
ID DRS1_YEAST
AC P32892;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Probable ATP-dependent RNA helicase DRS1.
GN DRS1 OR YIL008W OR L1345.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN 11
RP SEQUENCE FROM N.A.
RA MEDLINE=93087480; PubMed=1454790;
RA R1pmaster T.L., Vaughn G.P., Woolford J.L. Jr.;
RA "A putative ATP-dependent RNA helicase involved in Saccharomyces
RA cerevisiae ribosome assembly."

```

```

RL Proc. Natl. Acad. Sci. U.S.A. 89:11131-11135(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=S288c / F123;
RX MEDLINE=96405918; PubMed=8810043;
RA Miosga T., Zimmermann F.K.;
RT "Sequence analysis of the CEN12 region of Saccharomyces cerevisiae on
RT a 4.7 kb fragment of chromosome XII including an open reading frame
RT homologous to the human cyclic fibroblasts transmembrane conductance
RT regulator protein CFT2."
RL Yeast 12:693-708(1996).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=S288c / AB972;
RX MEDLINE=97131267; PubMed=9169871;
RA Johnston M., Hillier L., Riles L., Albermann K., Andre B., Ansgorge W.,
RA Benes V., Brueckner M., Deltis H., Dubois E., Dueserhoef A.,
RA Eutlian K.-D., Floeth M., Goffeau A., Hedling U., Heumann K.,
RA Heuss-Neitzel D., Hilbert H., Hilger F., Kleine K., Koetter P.,
RA Louis E.J., Messenguy F., Mewes H.-W., Miosga T., Mosecl D.,
RA Mueller-Auer S., Newbich U., Obermaier B., Pirzandi E., Pohl T.M.,
RA Portetelle D., Funtelle B., Reemann S., Rieger M., Rinke M., Rose M.,
RA Scharte M., Scherrens B., Scholler P., Schwager C., Schwarz S.,
RA Underwood A.P., Urrutiarazu L.A., Vandenbol M., Verhaaselt P.,
RA Vierendeel F., Voet M., Volckaert G., Voss H., Wamburt R., Wedler B.,
RA Wedler H., Zimmermann F.K., Zollner A., Hant J., Hobeisel J.D.;
RT "The nucleotide sequence of Saccharomyces cerevisiae chromosome XII."
RT Nature 387:87-90(1997).
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- SIMILARITY: BELONGS TO THE DEAD BOX HELICASE FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL L00683; AAA34666.1; -
DR EMBL X91488; CAA62783.1; -
DR EMBL Z73113; CAA97452.1; -
DR PIR S64750; S64750.
DR HSP; Q58083; 1HV8.
DR SGD; S0003931; DRS1.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR000629; DEAD box.
DR InterPro; IPR001650; Helicase_C.
DR Pfam; PF00270; DEAD; 1.
DR Pfam; PF00271; Helicase_C; 1.
DR SMART; SM00487; DEXDC; 1.
DR SMART; SM00490; HELIC; 1.
DR PROSITE; PS00039; DEAD ATP HELICASE; 1.
KM ATP-binding; RNA-binding; Helicase; Nuclear protein.
FT DOMAIN 170 190
FT NP_BIND 275 282
FT SITE 385 388
FT CONFLICT 1 68
FT FT
SQ SEQUENCE 752 AA; 84843 MW; 60747607AE58A48 CRC64;
Query Match
Best Local Similarity 21.8%; Score 85.5; DB 1; Length 752;
Matches 38; Conservative 29; Mismatches 64; Indels 43; Gaps 8;
QY 7 ILNKDTGEVSELKPRVTVYVTONGKEMSTIVSEBDFILPVYKGELEKGYQFD----- 59
DB 26 ILDSSDDEKVEAKK---TTKGRKGRK-NKKYVSEGNLDBDVHEDLDKFKFDLDADPT 81
QY 60 -----GMEI--SGFEGKQAGVYINLSKDTPIKPVFK-----IEKKEENKPTD 104

```

Db 82 SNFGNMFALGESKNDQAEFV--KQVDLDKIRRGGLVKNAMHISKQEEETEK-K 138

Qy 105 VSRRKDN-----PQVNHSQLNESHREKDLQREHSQKSDSTQVT 144

Db 139 VEKENDSDDEBLANDGFGMGAPMNNNGDENQSBEEBEEBKEBEEBEEBQEMT 192

RESULT 25

RGAL YEAST STANDARD; PRT; 1007 AA.

AC P39083; P39934; 01-FEB-1995 (Rel. 31, Created)

DT 01-FEB-1995 (Rel. 31, Last sequence update)

DT 15-SEP-2003 (Rel. 42, Last annotation update)

DE Rho-type GTPase-activating protein 1.

OS RGA1 OR DBM1 OR THB1 OR YOR127W OR O3290 OR YOR3290W.

OC Saccharomyces cerevisiae (Baker's yeast).

OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.

NCBI\_TaxID=4932;

RM NCB1

RN [1]

RP SEQUENCE FROM N.A., AND MUTAGENESIS.

RC STRAIN=S288C;

RX MEDLINE=96239492; PubMed=8657111;

RA Chen G.-C., Zheng L., Chan C.S.M.;

RT "The LIM domain-containing Dbm1 GTPase-activating protein is required for normal cellular morphogenesis in *Saccharomyces cerevisiae*."

RL Mol. Cell. Biol. 16:1376-1390(1996).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=S288C / FY1679;

RX MEDLINE=97060020; PubMed=8904341;

RA Mewam S., Rechmann S., Benes V., Voss H., Schwager C., Vleck C.,

RA Siegemann J., Zimmermann J., Rife H., Paces V., Ansoerge M.;

RT "Sequencing and analysis of 51 kb on the right arm of chromosome XV from *Saccharomyces cerevisiae* reveals 30 open reading frames."

RL Yeast 12:281-288(1996).

RN [3]

RP SEQUENCE FROM N.A.

RX MEDLINE=97344368; PubMed=9200815;

RA Voss H., Benes V., Andrade M.A., Valencia A., Rechmann S., Teodoru C.,

RA Schwager C., Paces V., Sander C., Ansoerge M.;

RT "DNA sequencing and analysis of 130 kb from yeast chromosome XV."

RL Yeast 13:655-672(1997).

RN [4]

RP SEQUENCE FROM N.A.

RC STRAIN=S288C;

RX MEDLINE=96101594; PubMed=7498791;

RA Stevenson B.J., Ferguson B., de Virgili C., Bi E., Pringle J.R.,

RA Amerer G., Sprague G.F. Jr.;

RT "Mutation of RGA1, which encodes a putative GTPase-activating protein for the polarity-establishment protein Cdc42p, activates the pheromone-response pathway in the yeast *Saccharomyces cerevisiae*."

RL Genes Dev. 9:2949-2963(1995).

RN [5]

RP SEQUENCE OF 570-639 FROM N.A.

RX STRAIN=SNY243;

RX MEDLINE=93087574; PubMed=1454852;

RA Ramer S.W., Ellledge S.J., Davis R.W.;

RT "Dominant genetics using a yeast genomic library under the control of a strong inducible promoter."

RL Proc. Natl. Acad. Sci. U.S.A. 89:11589-11593(1992).

CC -1- FUNCTION: GAPAS-ACTIVATING PROTEIN (GAP) FOR CDC42 AND/OR RHO1. NEGATIVE REGULATOR OF THE PHEROMONE-RESPONSE PATHWAY THROUGH THE STE20 PROTEIN KINASE. ACTS AT A STEP BETWEEN THE G-PROTEIN AND THE MAP KINASE MODULE. DOMINANT SUPPRESSOR OF BUD EMERGENCE DEFECT CAUSED BY DELETION OF IPL2/BEM2. INVOLVED IN THE CONTROL OF POLARIZED CELL GROWTH AND PROPER BUD SITE SELECTION.

CC -1- SIMILARITY: Contains 2 LIM zinc-binding domains.

CC -1- SIMILARITY: Contains 1 Rho-GAP domain.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation -

CC the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to [license@sib-ch](mailto:license@sib-ch)).

CC -----

DR EMBL; U07421; AAA16875.1; -

DR EMBL; X90518; CAA62108.1; -

DR EMBL; X94335; CAA64046.1; -

DR EMBL; 275035; CAA99326.1; -

DR EMBL; X90950; CAA62445.1; -

DR EMBL; L02617; AAA35153.1; -

DR PIR; S48535; S48535.

DR SGP; S0005653; RGA1.

DR GO; GO:0005100; F:Rho GTPase activator activity; IPI.

DR GO; GO:0004871; F:signal transducer activity; IPI.

DR GO; GO:0007015; P:actin filament organization; IMP.

DR GO; GO:0007118; P:apical bud growth; IPI.

DR GO; GO:0000283; P:establishment of cell polarity (sensu Sacch. . .); IPI.

DR GO; GO:0007125; P:invasive growth; IPI.

DR GO; GO:0007119; P:isotropic bud growth; IPI.

DR GO; GO:0007124; P:pseudohyphal growth; IPI.

DR GO; GO:0000750; P:signal transduction during conjugation with. . .; IGI.

DR GO; GO:0007264; P:small GTPase mediated signal transduction; IPI.

DR InterPro; IPR001781; LIM.

DR InterPro; IPR000198; RhoGAP.

DR Pfam; PF00412; LIM; 1.

DR Pfam; PF00620; RhoGAP; 1.

DR ProDom; PD000094; LIM; 2.

DR SMART; SM00324; RhoGAP; 1.

DR SMART; SM00324; RhoGAP; 1.

DR PROSITE; PS00478; LIM DOMAIN 1; 1.

DR PROSITE; PS50023; LIM DOMAIN 2; 2.

DR PROSITE; PS50238; RHO GAP; 1.

KW GTPase activation; Repeat; LIM domain; Metal-binding; Zinc; Pheromone response.

FT DOMAIN 13 66 LIM 1.

FT DOMAIN 70 122 LIM 2.

FT DOMAIN 791 1006 RHO-GAP.

FT VARIANT 866 866 V -> A.

FT VARIANT 898 898 K -> R.

FT VARIANT 926 926 S -> G.

FT MUTAGEN 40 40 C->S; BIPOLAR BUDDING.

FT MUTAGEN 37 37 C->S; BIPOLAR BUDDING.

FT MUTAGEN 98 98 C->S; BIPOLAR BUDDING.

FT MUTAGEN 101 101 C->S; BIPOLAR BUDDING.

FT CONFLICT 457 457 D -> E (IN REF. 4).

FT CONFLICT 507 507 T -> P (IN REF. 4).

SQ SEQUENCE 1007 AA; 112831 MW; C805411B57553791 CRC64;

Query Match 10.1%; Score 85.5; DB 1; Length 1007;

Best Local Similarity 21.0%; Pred. No. 59;

Matches 34; Conservative 34; Mismatches 69; Indels 25; Gaps 3;

Qy 11 DTGVSSELRVAVVTIONGKMSSTI-----VSBDFPILPVYKGLKKGQFGWBS 64

Db 349 MTGSIQMDPSLRKLVNNIYETNALQRPVEVVKEDRSVPDLGAGVQBOAEKYSYNN 408

Qy 65 GFEEKDAGVYINISKDTFFIKPVFKIEKKEENKPFVDSKKKDNQVNHSQLNBSHR 124

Db 409 SGKGRKTSRSLRSRSKDLMT-----NLKSRATGQSNVNLSPASRYTSRR 454

Qy 125 KEDQREHSHQSDSTQVATVLD-----KNNISKSTNN 161

Db 455 SQDLMRDNDSHGDTNRSNSTSIDLVNKKSLNRYKFTDN 496

RESULT 26

MABP RAT STANDARD; PRT; 2459 AA.

AC P15205; Q62958; Q9ER21; Q9QW92;

DT 01-APR-1990 (Rel. 14, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Microtubule-associated protein 1B (MAP 1B) (Neuraxin) [Contains: MAP1  
 light chain LC1].  
 GN MAP1B.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 CX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE OF 1-142 FROM N.A.  
 RC STRAIN=Sprague-Dawley; TISSUE=Testis;  
 RX MEDLINE=96257242; PubMed=8666295;  
 RA Liu D., Fischer I.;  
 RT "Isolation and sequencing of the 5' end of the rat microtubule-  
 associated protein (MAP1B)-encoding cDNA.";  
 RL Gene 172:307-308(1996).  
 RN [2]  
 RP SEQUENCE OF 96-2459 FROM N.A., DOMAIN, AND INDUCTION.  
 RC STRAIN=Sprague-Dawley; TISSUE=Brain, and Gliat tumor;  
 RX MEDLINE=92347374; PubMed=1639092;  
 RA Zauner W., Kratz J., Staunton J., Feick P., Wiche G.;  
 RT "Identification of two distinct microtubule binding domains on  
 recombinant rat MAP 1B.";  
 RL Eur. J. Cell Biol. 57:66-74(1992).  
 RN [3]  
 RP SEQUENCE OF 1541-2459 FROM N.A., AND TISSUE SPECIFICITY.  
 RC TISSUE=Spinal cord;  
 RX MEDLINE=90059871; PubMed=2555150;  
 RA Rientz A., Gremingloh G., Hermans-Borgmeyer I., Kirsch J.,  
 RA Litaauer U.Z., Prior P., Gundelfinger E.D., Schmitt B., Betz H.;  
 RT "Neuraxin, a novel putative structural protein of the rat central  
 nervous system that is immunologically related to microtubule-  
 associated protein 5.";  
 RL EMBO J. 8:2879-2888(1989).  
 RN [4]  
 RP DEVELOPMENTAL STAGE, AND PHOSPHORYLATION.  
 RX MEDLINE=97405699; PubMed=9260743;  
 RA Ma D., Nothias F., Boyne L.J., Fischer I.;  
 RT "Differential regulation of microtubule-associated protein 1B (MAP1B)  
 in rat CNS and PNS during development.";  
 RL J. Neurosci. Res. 49:319-332(1997).  
 CC -!- FUNCTION: The function of brain MAPs is essentially unknown.  
 CC Phosphorylated MAP1B may play a role in the cytoskeletal changes  
 CC that accompany neurite extension. Possibly MAP1B binds to at least  
 CC two tubulin subunits in the polymer, and this bridging of subunits  
 CC might be involved in nucleating microtubule polymerization and in  
 CC stabilizing microtubules.  
 CC -!- SUBUNIT: 3 different light chains, LC1, LC2 and LC3, can associate  
 CC with MAP1A and MAP1B proteins.  
 CC -!- TISSUE SPECIFICITY: Nervous system (spinal cord, brain stem,  
 CC cerebellum and cerebrum). Not expressed in liver, spleen, kidney,  
 CC heart or muscle.  
 CC -!- DEVELOPMENTAL STAGE: In cerebral cortex, spinal cord and sciatic  
 CC nerve levels are high early in development but decrease during  
 CC postnatal development and are low in adults. In dorsal root  
 CC ganglia levels remain high throughout development.  
 CC -!- INDUCTION: By nerve growth factor.  
 CC -!- DOMAIN: Has a highly basic region with many copies of the sequence  
 CC KKEE and KKEE/V, repeated but not at fixed intervals, which is  
 CC responsible for the binding of MAP1B to microtubules.  
 CC -!- PTM: LC1 is coexpressed with MAP1B. It is a polypeptide generated  
 CC from MAP1B by proteolytic processing. It is free to associate with  
 CC both MAP1A and MAP1B. It interacts with the amino-terminal region  
 CC of MAP1B (by similarity).  
 CC -!- PTM: Phosphorylated.  
 CC -!- SIMILARITY: TO MAP1A.  
 CC -!- CAUTION: A C-terminal fragment of this protein (residues 1597 to  
 CC 2459) was originally described as neuraxin in Ref.3.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: U52950; AAB17068.1; -  
 CC DR EMBL: X60370; CAC16162.1; -  
 CC DR EMBL: X16623; CAA34620.1; ALU\_SEQ.  
 CC DR PIR: A56577; A56577.  
 CC DR InterPro: IPR000102; MAP1B\_neuraxin.  
 CC DR Pfam: PF00414; MAP1B\_neuraxin; 10.  
 CC DR PROSITE: PS00230; MAP1B\_NEURAXIN; 8.  
 CC KW Microtubules; Repeat; Phosphorylation.  
 CC CHAIN ? 2459 MAP1 LIGHT CHAIN LC1.  
 CC FT REPEAT 1869 1885 MAP1B 1.  
 CC FT REPEAT 1886 1902 MAP1B 2.  
 CC FT REPEAT 1903 1919 MAP1B 3.  
 CC FT REPEAT 1920 1936 MAP1B 4.  
 CC FT REPEAT 1937 1953 MAP1B 5.  
 CC FT REPEAT 1954 1970 MAP1B 6.  
 CC FT REPEAT 1988 2004 MAP1B 7.  
 CC FT REPEAT 2005 2021 MAP1B 8.  
 CC FT REPEAT 2022 2038 MAP1B 9.  
 CC FT REPEAT 2039 2055 MAP1B 10.  
 CC FT REPEAT 559 1035 MAP1B 10.  
 CC FT DOMAIN 588 786 LYS-RICH (HIGHLY BASIC, CONTAINS MANY  
 CC KKEE AND KKEE/V REPEATS).  
 CC FT DOMAIN 2224 2312 LYS-RICH.  
 CC FT CONFLICT 127 127 M -> V (IN REF. 1).  
 CC FT CONFLICT 140 140 T -> S (IN REF. 1).  
 CC FT CONFLICT 2112 2112 R -> K (IN REF. 3).  
 CC FT CONFLICT 2169 2169 L -> I (IN REF. 3).  
 CC SQ SEQUENCE 2459 AA; 269497 MW; 2E3F6872DEDB8BA2 CRC64;  
 CC -----  
 CC Query Match 10.1%; Score 85.5; DB 1; Length 2459;  
 CC Best Local Similarity 24.4%; Pred. No. 1.5e+02;  
 CC Matches 40; Conservative 26; Mismatches 53; Indels 45; Gaps 8;  
 CC -----  
 CC QY 4 KEFLNKDTGEVSELPKPRVLTIONGKMSSTVSEBDFILPVYKGLSEKGYQPGWMI 63  
 CC DB 583 EKVIYKDKPKGVESKPSVTEKEVPSKEQS-----PV-KAEVA-----EK 622  
 CC QY 64 SGPEGKQAGVYINISKDTFIYVFK-KLEKKEENKTPDVSKKKDPVNNHSQLNS 122  
 CC DB 623 AATESKP-----KTKKRVVKEEIKTPEEKKEE--XPKKVVAKED----- 662  
 CC QY 123 HKKEPLQREHSHQSDSTKDVATVL--DKNNISSKSTNNPNK 164  
 CC DB 663 --KTEPLKDEKPKGEARKEIKKEKKEKELKXVKKETEPLK 704  
 CC -----  
 CC RESULT 27  
 CC YALE SCHPO STANDARD; PRT; 719 AA.  
 CC AC Q10342; Q9URU8;  
 CC DT 01-OCT-1996 (Rel. 34, Created)  
 CC DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 CC DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 CC DR Hypothetical protein C106.14c in chromosome II.  
 CC GN SPBC106.14c.  
 CC OS Schizosaccharomyces pombe (Fission Yeast).  
 CC OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
 CC OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
 CC OC Schizosaccharomycetes.  
 CC OX NCBI\_TaxID=4896;  
 CC RN [1]  
 CC RP SEQUENCE FROM N.A.  
 CC RC STRAIN=972;  
 CC RX MEDLINE=21848401; PubMed=11859360;  
 CC Wood V., Williams R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,  
 CC Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,  
 CC Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,  
 CC Collins M., Connor R., Cronin A., Davis P., Feltham T., Fraser A.,  
 CC Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,

```

RA Chen Z., Wen Y. ;
RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
CC
CC -!- FUNCTION: One of the essential components for the initiation of
CC protein synthesis. Protects formylmethionyl-tRNA from spontaneous
CC hydrolysis and promotes its binding to the 30S ribosomal subunits.
CC Also involved in the hydrolysis of GTP during the formation of the
CC 70S ribosomal complex (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: BELONGS TO THE IF-2 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE016747; AAC04542.1; .
DR HAMAP; MP_001007; -; 1.
DR Pfam; PF00009; GTP_EFTU; 1.
DR Pfam; PF03144; DR_PFTU_D2; 1.
DR Pfam; PF04760; IF2_N; 2.
DR ProDom; PD186100; IF2; 1.
DR TIGRFAMs; TIGR00487; IF-2; 1.
DR TIGRFAMs; TIGR00231; small_GTP; 1.
DR PROSITE; PS01176; IF2; 1.
DR Initiation factor; Protein biosynthesis; GTP-binding;
KM Complete proteome.
FT DOMAIN 225 373 G-DOMAIN.
FT NP_BIND 231 238 GTP (BY SIMILARITY).
FT NP_BIND 277 281 GTP (BY SIMILARITY).
FT NP_BIND 331 334 GTP (BY SIMILARITY).
SQ
SEQUENCE 720 AA; 79343 MW; 07FB5A6A59CF970C CRC64;

Query March 10.0%; Score 85; DB 1; Length 720;
Best Local Similarity 28.0%; Pred. No. 45;
Matches 23; Conservative 15; Mismatches 36; Indels 8; Gaps 2;

QY 91 IBEKKEENKPTFVSKKKD---NPQVNSQLNESHREKLEQREHSOKSDYK---D 142
Db 38 LEEGQIALDKKFRASAKDTKNTKNTQNTQKSNKNSNDKRGQSKNSKPTKKKEQN 97
QY 143 VTATVLDKNNISSKSTNNPNK 164
Db 98 MKGKQNKNNKTKNTKNTKNTKNTK 119

RESULT 29
Y040 MYCPN STANDARD; PRT; 657 AA.
AC P75062;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical lipidoprotein MG040 homolog precursor (D09_orf657).
GN MF0053 OR MF102.
OS Mycoplasma pneumoniae.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxId=2104;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 29342 / M129;
RX MEDLINE=97105885; Pubmed=8948633;
RA Himmelfreich R., Hilbert H., Plagens H., Pirkl E., Li B.-C.,
RA Hermann R.
RT "Complete sequence analysis of the genome of the bacterium Mycoplasma
RT pneumoniae."
RL Nucleic Acids Res. 24:4420-4449 (1996).
CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a lipid anchor
CC (Potential).
CC -!- SIMILARITY: SOME, TO T.PALLIDIUM TWPC.
CC -----

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
 CC EMBL: AE000012; AAB95750.1; -.  
 CC PIR: S73428; S73428.  
 CC InterPro: IPR003760; Bmp.  
 CC Pfam: PF02608; Bmp.1.  
 CC PROSITE: PS00013; PROKAR\_LIPOPROTEIN\_1.  
 CC Hypothetical protein; Lipoprotein; Membrane; Signal;  
 CC Complete proteome.  
 CC SIGNAL 1 26 POTENTIAL.  
 CC CHAIN 27 657 HYPOTHEMETICAL LIPOPROTEIN MG040 HOMOLOG.  
 CC LIPID 27 27 N-ACTIV DIGLYCERIDE (POTENTIAL).  
 CC FT SEQUENCE 657 AA; 71671 MW; 6147B1A5673606A9 CRC64;  
 CC SQ

Query Match 10.0%; Score 84.5; DB 1; Length 657;  
 Best Local Similarity 16.7%; Pred. No. 44;  
 Matches 42; Conservative 36; Mismatches 78; Indels 95; Gaps 5;

QY 8 LNKDTGEVSELSKPHRYVTITIOGKEM-----SRTIVSE 41  
 Db 343 VNLAVNEVARRKRAHTAIIIVDSAGSLIDINQAPDKOLIKNKKIIPSSIKALDAVLE 402  
 QY 42 DFILPVYKELGKGYQFDG-----WEISGEKKKADY 74  
 Db 403 NMLIAIQGSDNNQYKGPYNNIGVTGSSVGISEAGYFLIDPVWKTQOGKSMANN 462  
 QY 75 VINL-----SKDTFKPVFKIEKKE----- 96  
 Db 463 MTNLKRLSSDDPTNTKKALKKSVTHKNGSDKDGIIIGKYSMLTKKSTPTVAVAKSWTDNN 522  
 QY 97 ---EENKPLFDV-SKKKNPQVNHQSLNESHRKEDLQREKHSOKSSTQVATATLDKRN 152  
 Db 523 SGTBEKKNLSEVDTKKKEKESKGTQSGDQNGKGTNDIISKYSKLTMTTMMNKV 582  
 QY 153 ISSKSTNNPN 163  
 Db 583 MSSKKQSSDN 593

RESULT 30  
 IGA0\_HAETN STANDARD; PRT; 1694 AA.  
 AC P44969;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DT Immunoglobulin A1 protease precursor (EC 3.4.21.72) (IGA1 protease).  
 DR IGA OR IGA1 OR HI0990.  
 OS Haemophilus influenzae.  
 CC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;  
 CC Pasteurellaceae; Haemophilus.  
 CC NCBI\_TaxID=727;  
 OC [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=serotype D;  
 RL Wright A., Fishman Y., Tai F., Plaut A.G.,  
 RL Submitted (May-1991) to the EMBL/Genbank/DBJ databases.  
 RP [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=RD / KW20 / ATCC 51907;  
 RX MEDLINE=95350630; PubMed=7542800;  
 RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,  
 RA Kevlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,  
 RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,  
 RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,  
 RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,  
 RA Uterback T.R., Hanna M.C., Nguyen D.T., Saudak D.M., Brandon R.C.,

RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,  
 RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,  
 RA Venter J.C.;  
 RT "Whole-genome random sequencing and assembly of Haemophilus influenzae  
 RT Rd.";  
 RL Science 269:496-512(1995).  
 CC -!- FUNCTION: VIRULENCE FACTOR: CLEAVES HOST IMMUNOGLOBULIN A  
 CC PRODUCING INTACT FC AND FAB FRAGMENTS.  
 CC -!- CATALYTIC ACTIVITY: Cleavage of immunoglobulin A molecules at  
 CC certain Pro-I-Xaa bonds in the hinge region. No small molecule  
 CC substrates are known.  
 CC -!- SUBCELLULAR LOCATION: Secreted.  
 CC -!- DOMAIN: THE SIGNAL PEPTIDE GUIDES THE PRECURSOR TO THE PERIPLASMIC  
 CC SPACE, AND THE CARBOXY-TERMINAL HELPER DOMAIN ASSOCIATES WITH THE  
 CC OUTER MEMBRANE TO FORM A PORE FOR EXCRETION OF THE PROTEASE  
 CC DOMAIN. THE HELPER DOMAIN IS THEN RELEASED BY AUTOPROTEOLYSIS (BY  
 CC SIMILARITY).  
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S6.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
 CC EMBL: X59800; -; NOT ANNOTATED\_CDS.  
 CC EMBL: U32779; AAC22651.1; -.  
 CC PIR: H64106; H64106.  
 CC MEROPS: S06.001; -.  
 CC TIGR: H10990; -.  
 CC InterPro: IPR006315; Autotransport.  
 CC InterPro: IPR005546; Autotransporter.  
 CC InterPro: IPR000710; IGA S6.  
 CC InterPro: IPR004899; Pertactin.  
 CC Pfam: PF03797; Autotransporter; 1.  
 CC Pfam: PF02395; IGA1; 1.  
 CC Pfam: PF03212; Pertactin; 1.  
 CC PRINTS: PR00921; IGASERPTASE.  
 CC TIGRFAMs: TIGR01414; autotrans barl; 1.  
 CC Hydroxylase; Serine protease; Transmembrane; Zymogen; Signal;  
 CC Complete proteome.  
 CC FT SIGNAL 1 25 POTENTIAL.  
 CC FT CHAIN 26 1014 IMMUNOGLOBULIN A1 PROTEASE.  
 CC FT PROPEP 1015 1694 HELPER PEPTIDE (POTENTIAL).  
 CC FT ACT\_SITE 288 288 PROBABLE.  
 CC FT CONFLICT 253 254 EN -> GV (IN REF. 1).  
 CC FT CONFLICT 272 272 G -> A (IN REF. 1).  
 CC FT CONFLICT 464 464 G -> E (IN REF. 1).  
 CC FT CONFLICT 866 866 S -> T (IN REF. 1).  
 CC FT CONFLICT 1036 1036 A -> D (IN REF. 1).  
 CC FT CONFLICT 1074 1074 A -> G (IN REF. 1).  
 CC FT CONFLICT 1421 1421 A -> G (IN REF. 1).  
 CC FT CONFLICT 1545 1545 H -> T (IN REF. 1).  
 CC FT SEQUENCE 1694 AA; 185539 MW; C52427013P931178C CRC64;  
 CC SQ

Query Match 10.0%; Score 84.5; DB 1; Length 1694;  
 Best Local Similarity 26.1%; Pred. No. 1.2e+02;  
 Matches 24; Conservative 15; Mismatches 50; Indels 3; Gaps 1;

QY 76 INLSKDTPIKPVFKIEKKEENKPTDVSKKKNPQVNHQSLNESHRKEDLQREKHS- 134  
 Db 1288 INTGSAFATITETAEKSDKPKQETETASTEDASQHKANTAVANSSSDPKSRRRRSI 1347

QY 135 --QKSDTKQVATATLDKNNISSKSTNNPK 164  
 Db 1348 SQPQETSAERTTAASTDETTIADNSKRSKSPNR 1379

RESULT 31  
 HGV2\_HALRO STANDARD; PRT; 510 AA.  
 ID HGV2\_HALRO

```

AC 002508;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Protein HGV2.
GN HGV2.
OS Halocynthia roretzi (Sea squirt).
OC Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea;
OC Stolidobranchia; Pyruidae; Halocynthia.
OC NCBI_TaxID=7729;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Oocyte;
RX MEDLINE=93224498; PubMed=8468323;
RA Fujiwara S., Kawahara H., Makabe K.W., Satoh N.;
RT "A complementary DNA for an ascidian embryonic nuclear antigen Hgv2
RT encodes a protein closely related to the amphibian histone-binding
RT protein N1."
RL Cell 113:189-195(1993).
CC -1- FUNCTION: MAY FUNCTION AS A NUCLEOSOME ASSEMBLY FACTOR DURING
CC -1- RAPID EMBRYONIC CELL DIVISIONS.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- TISSUE SPECIFICITY: EMBRYO AND LARVAE.
CC -1- DEVELOPMENTAL STAGE: THIS PROTEIN IS DETECTED IN THE NUCLEI
CC OF ALL CELLS IN EMBRYOS AND LARVAE BUT IS NOT DETECTED IN THE
CC CELLS OF METAMORPHOSED JUVENILES.
CC -1- SIMILARITY: TO XENOPUS LAEVIS HISTONE BINDING PROTEIN N1.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D13541; BA02741.1; -.
DR PIR; JX0254; JX0254.
DR InterPro; IPR001440; TPR.
DR Pfam; PF00515; TPR; 2.
KW Nuclear protein; Developmental protein.
FT DOMAIN 98 143 ASP/GLU-RICH (ACIDIC).
FT DOMAIN 219 228 ASP/GLU-RICH (ACIDIC).
FT DOMAIN 444 451 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT DOMAIN 465 471 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
SQ SEQUENCE 510 AA; 56871 MW; D9961B5953B976PF CRC64;

Query Match 9.9%; Score 84; DB 1; Length 510;
Best Local Similarity 28.4%; Pred. No. 36;
Matches 50; Conservative 25; Mismatches 61; Indels 40; Gaps 12;

QY 4 KEPIIN-KDTGEVSELKPHRYVTVTIIONGKMSSTIVSEBDPILPYKGELEKGYOPDGM 62
DB 354 KEIISECKEVELEKELIP-----DINSKI--EDVILA--KKOMQK--LDG-- 392
QY 63 ISGF-----EGKKGAGVINTLSKOTFIK--PVFK---KIEEKEENKPTPVSKKKNP 112
DB 393 -SPFQASKEGSSSGGLGASTSDKPCSTIPTRKAPTSPVPAKSPSDITHLVREKRSP 451
QY 113 -----QVNSQINSHRKKEDLQREHSQKSDSTKQVATVLDKNNISSTNNPK 164
DB 452 DEDNQPAESKESKKAQKEETBEATNGSHSAVKQT-DVTDKNGTNGHSGT--PKK 504

RESULT 32
RBP2_PLAVB STANDARD; PRT; 1251 AA.
ID_RBP2_PLAVB
AC Q00793;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Reticulocyte binding protein 2 (Fragment).
RN RBP2.

```

```

OS Plasmodium vivax (strain Belem).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OC NCBI_TaxID=31273;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92315338; PubMed=1617731;
RA Galinski M.R., Medina C.C., Ingravallo P., Barnwell J.W.;
RT "A reticulocyte-binding protein complex of Plasmodium vivax
RT merozoites."
RL Cell 69:1213-1226(1992).
CC -1- FUNCTION: INVOLVED IN RETICULOCYTE ADHESION. SPECIFICALLY BINDS TO
CC -1- HUMAN RETICULOCYTE CELLS.
CC -1- SUBCELLULAR LOCATION: Membrane-bound (Probable).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M88098; AAA29744.1; -.
KW Malaria; Receptor; Membrane.
FT NON TER 1
FT NON TER 1251
SQ SEQUENCE 1251 AA; 143741 MW; 54BA51C7404AC572 CRC64;

Query Match 9.9%; Score 84; DB 1; Length 1251;
Best Local Similarity 23.0%; Pred. No. 96;
Matches 42; Conservative 37; Mismatches 72; Indels 32; Gaps 9;

QY 4 KEPIIN-KD-TGEVSELKPH--RYVTVTIIONGKMSSTI-----VSEBDPILPYKGE 52
DB 253 REIMAKKDBIKSYLSEIKYKDKCTTEISNGRGKDIKLEFKFKPNBSNSNKNVINEI 312
QY 53 EKGYPDGMELSGFR-GKKDAGVINTL--SKOTFIKPVFKK-----IEEKEENKPTD 104
DB 313 MENINSEBYLKDIDDAKQASTKVELFKHETTSINIKSEELIGVETRSOKINKAD 372
QY 105 VSK--KKNPQV-----NHSQINSHRKKEDLQREHSQKSDSTKQVATVLD--KN 151
DB 373 IMKEIERHNSLQYQVKGFRQENLKNLBNPHNYDAEDELNDKSTNAKVLITNLESVGH 432
QY 152 NIS 154
DB 433 NIS 435

RESULT 33
PERR_PEA STANDARD; PRT; 360 AA.
ID_PERR_PEA
AC P10933;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Ferredoxin-NADP reductase, leaf isozyme, chloroplast precursor
DE (EC 1.18.1.2) (FNR).
GN PETH.
OS Pisum sativum (Garden pea).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids I; Fabales; Fabaceae; Papilionoidae; Viciae; Pisum.
OC NCBI_TaxID=3888;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Little Marvel; TISSUE=leaf;
RA Newman B.J., Gray J.C.;
RT "Characterisation of a full-length cDNA clone for pea ferredoxin-NADP+
RT reductase."
RL Plant Mol. Biol. 10:511-520(1988).
RN [2]
SQ SEQUENCE OF 270-360 FROM N.A., AND MUTAGENESIS.

```

RX MEDLINE=93374905; PubMed=8366077;  
 RA Orellano E.G., Calcaterra N.B., Carrillo N., Ceccarelli E.A.;  
 RT "Probing the role of the carboxyl-terminal region of ferredoxin-NADP+  
 reductase by site-directed mutagenesis and deletion analysis.";   
 RL J. Biol. Chem. 268:19267-19273(1993).  
 RN [3]  
 RP X-RAY CRYSTALLOGRAPHY (1.70 ANGSTROMS).  
 RX MEDLINE=93936739; PubMed=10467097;  
 RA Deng Z., Aliverti A., Zanetti G., Arakaki A.K., Oltado J.,  
 RA Orellano E.G., Calcaterra N.B., Ceccarelli E.A., Carrillo N.,  
 RA Karpus P.A.;  
 RT "A productive NADP+ binding mode of ferredoxin-NADP+ reductase  
 RT revealed by protein engineering and crystallographic studies.";   
 RL Nat. Struct. Biol. 6:847-853(1999).  
 CC -1- FUNCTION: MAY PLAY A KEY ROLE IN REGULATING THE RELATIVE AMOUNTS  
 CC OF CYCLIC AND NON-CYCLIC ELECTRON FLOW TO MEET THE DEMANDS OF THE  
 CC PLANT FOR ATP AND REDUCING POWER.  
 CC -1- CATALYTIC ACTIVITY: Reduced ferredoxin + NADP(+) = oxidized  
 CC ferredoxin + NADPH.  
 CC -1- COFACTOR: FAD.  
 CC -1- PATHWAY: FINAL STEP IN LINEAR PHOTOSYNTHETIC ELECTRON TRANSPORT  
 CC CHAIN; IT HAS ALSO BEEN IMPLICATED IN CYCLIC ELECTRON FLOW AROUND  
 CC PHOTOSYSTEM I, ITS ROLE BEING TO RETURN ELECTRONS FROM FERREDOXIN  
 CC TO THE CYTOCHROME B-F COMPLEX.  
 CC -1- SUBUNIT: Monomer.  
 CC -1- SUBCELLULAR LOCATION: CHLOROPLAST, STROMAL SIDE OF THE THYLAKOID  
 CC MEMBRANE IN THE VICINITY OF THE PHOTOSYSTEM I IN THE NON-STACKED  
 CC AND FRINGE PORTION OF THE MEMBRANE.  
 CC -1- MISCELLANEOUS: FNR IS PROBABLY ATTACHED TO THE MEMBRANE BY A  
 CC SPECIFIC BINDING PROTEIN.  
 CC -1- SIMILARITY: WITH OTHER SPECIES FNR.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: X12446; CAJ30978.1; -;  
 DR EMBL: I15565; AAB59349.1; -;  
 DR EMBL: I15567; AAB59303.1; -;  
 DR EMBL: I15569; AAB59304.1; -;  
 DR PIR: S04030; S04030.  
 DR PDB: 1QFZ; 01-SEP-99.  
 DR PDB: 1QFY; 01-SEP-99.  
 DR PDB: 1QGO; 01-SEP-99.  
 DR PDB: 1QGA; 01-SEP-99.  
 DR InterPro: IPR001709; FPN\_cyt\_reductase.  
 DR InterPro: IPR001433; Oxired\_FAD/NAD(P).  
 DR Pfam: PF00175; NAD\_binding\_1; 1.  
 DR PRINTS: PR00371; PFNCR.  
 KW Oxidoreductase; Flavo-protein; NADP; FAD; Chloroplast; Transit peptide;  
 KW Electron transport; Photosynthesis; Thylakoid; Membrane;  
 KW Multigene family; 3D-structure.  
 FT TRANSIT 1 52 CHLOROPLAST.  
 FT CHAIN 53 360 FERREDOXIN--NADP REDUCTASE, LEAF ISOZYME.  
 FT DISULFID 178 183 BY SIMILARITY.  
 FT NP\_BIND 212 230 NADP (RIBOSE PART) (BY SIMILARITY).  
 FT MUTAGEN 360 360 Y->W,F,S.  
 FT TURN 71 72  
 FT STRAND 78 78  
 FT TURN 80 81  
 FT STRAND 84 83  
 FT TURN 97 98  
 FT STRAND 103 109  
 FT TURN 111 112  
 FT STRAND 118 119  
 FT TURN 121 125  
 FT STRAND 129 132  
 FT TURN 131 132

FT STRAND 135 135  
 FT STRAND 139 143  
 FT TURN 147 148  
 FT TURN 150 151  
 FT STRAND 156 162  
 FT STRAND 165 167  
 FT TURN 169 170  
 FT STRAND 173 175  
 FT HELIX 177 184  
 FT TURN 187 188  
 FT STRAND 190 197  
 FT TURN 200 201  
 FT STRAND 204 204  
 FT TURN 207 208  
 FT STRAND 210 216  
 FT HELIX 217 220  
 FT HELIX 221 231  
 FT TURN 232 232  
 FT STRAND 243 250  
 FT HELIX 253 255  
 FT TURN 257 258  
 FT HELIX 259 268  
 FT TURN 270 272  
 FT STRAND 273 279  
 FT TURN 280 282  
 FT STRAND 284 284  
 FT TURN 286 287  
 FT STRAND 290 290  
 FT HELIX 293 297  
 FT HELIX 298 300  
 FT HELIX 301 309  
 FT TURN 311 312  
 FT STRAND 313 319  
 FT TURN 321 322  
 FT HELIX 323 337  
 FT TURN 338 339  
 FT HELIX 342 351  
 FT TURN 352 353  
 FT STRAND 355 360  
 SQ SEQUENCE 360 AA; 40194 MW; 7F1C10DBBA7B24 CRC64;  
 Query Match 9.8%; Score 83.5; DB 1; Length 360;  
 Best Local Similarity 24.7%; Pred. No. 27;  
 Matches 38; Conservative 19; Mismatches 50; Indels 47; Gaps 7;  
 QY 3 VKEFILNKDTGYEV-----SELRP-HRYVTYIIONGKEM-----SSTI 37  
 DB 161 VKRLVYTNDAIGHVYKGVCSNPLCDLKGSEVKTGPGVKKMLPKDPNATVIMLGCTGTGI 220  
 QY 38 VSEBDFILPVYKGELEKGYQFG--WEISGFGSKDAGVINLSKDTPIKPVKIEKK 95  
 DB 221 AFRSFLMTGF-FEKRDYQFNGLAWLPLG-----VPTSSLLYKGFBEKKMEKA 269  
 QY 96 EEEKNPTFVSKKKDKPNQVNSQLNESHREKDLQ 129  
 DB 270 PENFRLDPAVSRK-----QVNDKGRKMTIQ 294  
 RESULT 34  
 UZRL\_HUMAN  
 ID UZRL\_HUMAN STANDARD; PRT; 479 AA.  
 AC Q15695; Q1570;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE U2 small nuclear ribonucleoprotein auxiliary factor 35 kDa subunit  
 DE related-protein 1.  
 GN UZAF1-RS1 OR UZAF1RS1 OR UZAFBPL.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]



```

RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=96163878; PubMed=8586425;
RA Kitzgawa K., Wang X., Hatada I., Yamaoka T., Nojima H.,
RA Inazawa J., Aoe T., Mitsuya K., Oshimura M., Murata A., Monden M.,
RA Mukai T.;
RT "Isolation and mapping of human homologues of an imprinted mouse gene
RT U2af1-rel.";
RL Genomics 30:257-263(1995).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=96212931; PubMed=8630064;
RA Pearbairl R.S., Shibata H., Brozowska A., Yoshino K., Okuda K.,
RA Dejong P.J., Phase C., Chapman V.M., Hayashizaki Y., Held W.A.;
RT "Absence of imprinting in U2AFBPL, a human homologue of the imprinted
RT mouse gene U2afbp-rs.";
RL Biochem. Biophys. Res. Commun. 222:171-177(1996).
CC -1- SUBCELLULAR LOCATION: Nuclear (Potential).
CC -1- SIMILARITY: Contains 1 RNA recognition motif (RRM) domain.
CC -1- SIMILARITY: Contains 2 C3H1-type zinc fingers.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL, D49676; BAA08532.1; -
DR EMBL, U51224; AAA98669.1; -
DR MIM, 601079; -
DR GO, GO:0005634; C:nucleus; NAS.
DR GO, GO:0003723; F:RNA binding activity; NAS.
DR InterPro, IPR000504; RNA rec mot.
DR InterPro, IPR000571; znf_CCCH.
DR Pfam, PF00076; rrm_1.
DR Pfam, PF00642; zf-CCCH; 2.
DR SMART, SM00360; RRM; 1.
DR SMART, SM00356; Znf_C3H1; 2.
DR PROSITE, PS0102; RRM; 1.
DR PROSITE, PS00030; RRM_RNP_1; FALSE_NEG.
DR Nuclear protein; RNA-binding; Ribonucleoprotein; Zinc-finger; Repeat.
KM DNA-BINDING (RRM).
FT DOMAIN 203 309
SQ SEQUENCE 479 AA; 57643 MW; 96F326694BD4E7C0 CRC64;

Query Match 9.8%; Score 83.5; DB 1; Length 479;
Best Local Similarity 21.7%; Pred. No. 37;
Matches 26; Conservative 25; Mismatches 40; Indels 29; Gaps 4;

QY 70 KDAGVINLSKDTFKPVKPKIEKK-----EENKPTFDVSKKKDNQ 113
DB 44 RDSGLSQEEDFTLIE--QOLEEKLTERERRLHEETLREQKQBEFRKKEKEEA 101
QY 114 VN-----HSQINSHRKEDLQREHSOKSDTKDVTATVLD--KNINISKSTNNP 162
DB 102 KKWLEPQERKLKEQWKEQQRKEEREBEQKQEKKEKEBEAVQKLDQAEVDLKRSTWNP 161

RESULT 35
GAP PLAAF STANDARD; PRT; 678 AA.
AC P13816;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 15-JUL-1999 (Rel. 36, Last annotation update)
DE Glutamic acid-rich protein precursor.
GN GARP.
OS Plasmodium falciparum (Isolate FC27 / Papua New Guinea).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5837;
RN [1]
RP SEQUENCE FROM N.A.

```

```

RX MEDLINE=89040048; PubMed=2903445;
RA Triglia T., Stahl H.-D., Crewther P.E., Silva A., Anders R.F.,
RA Kemp D.J.;
RT "Structure of a Plasmodium falciparum gene that encodes a glutamic
RT acid-rich protein (GARP).";
RL Mol. Biochem. Parasitol. 31:199-202(1988).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL, J03998; AAA29605.1; -
DR PIR, A54514; A54514.
KM Repeat; Malaria; Antigen; Signal.
FT SIGNAL 1 25
FT CHAIN 26 678
FT DOMAIN 120 164 15 X 3 AA TANDEN REPEATS OF K-K-X.
FT DOMAIN 372 416 9 X APPROXIMATE TANDEN REPEATS.
FT DOMAIN 417 441 5 X APPROXIMATE TANDEN REPEATS.
FT DOMAIN 576 604 POLY-GLU.
FT DOMAIN 605 653 7 X APPROXIMATE TANDEN REPEATS.
FT DOMAIN 654 663 POLY-GLU.
SQ SEQUENCE 678 AA; 80551 MW; 2A8F85606496BA9E CRC64;

Query Match 9.8%; Score 83.5; DB 1; Length 678;
Best Local Similarity 24.6%; Pred. No. 54;
Matches 44; Conservative 33; Mismatches 65; Indels 37; Gaps 10;

QY 7 ILNKD---TGEV---SEIKPRVVTYTONGKMSSTIYSE---DFILPVKGLLEKG 55
DB 37 ILNKSPDSITGRILNETLEKKN-----DNKSSETLKKEKRDVDVPTSDNMLKNA 90
QY 56 YQPDGHEIGSPREKGDAGVINLSKDTFKPVKPKIEKK-----EENKPTFDVSKKK 109
DB 91 HNNN--EIS---SGDPTNINVDKNENSVKDKKQKKKKKKKKKKKKKKKKKKKK 145
QY 110 DNPQVHQSQINSHRKEDLQREHSO-----XSDTKDVTATVLDKNINISK--STNN 161
DB 146 DKKEKKKKK--EKKKKKKKKKKENSEVSLYKTGQHKPKQKATHEGSENDREBVSSEINN 202

RESULT 36
ID HS9A HUMAN STANDARD; PRT; 731 AA.
AC P07900; Q9BVO5;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Heat shock protein HSP 90-alpha (HSP 86).
GN HSPCA OR HSPCL OR HSP90A.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Peripheral blood lymphocytes;
RX MEDLINE=89386066; PubMed=2780322;
RA Soeda E., Yokoyama K., Yamazaki M., Akaogi K., Miwa T., Inai T.;
RT "Nucleotide sequence of a full-length cDNA for 90 kDa heat-shock
RT protein from human peripheral blood lymphocytes.";
RL Nucleic Acids Res. 17:7108-7108(1989).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=91242090; PubMed=1368637;
RA Yamazaki M., Tashiro H., Yokoyama K., Soeda E.;
RT "Molecular cloning of cDNA encoding a human heat-shock protein whose
RT expression is induced by adenovirus type 12 B1A in HeLa cells.";
RL Agric. Biol. Chem. 54:3163-3170(1990).

```

RN [3] SEQUENCE FROM N.A.  
 RP TISSUE=Placenta;  
 RC MEDLINE=89363979; PubMed=2527334;  
 RA Hickey B., Brandon S.E., Smaile G., Lloyd D., Weber L.A.;  
 RT "Sequence and regulation of a gene encoding a human 89-kilodalton  
 RT heat shock protein.";  
 RL Mol. Cell. Biol. 9:2615-2626(1989).  
 RN [4] SEQUENCE OF 1-311 FROM N.A.  
 RP MEDLINE=8056312; PubMed=2445630;  
 RA Hoffmann T., Hovemann B.;  
 RT "Cloning and nucleotide sequence of the murine hsp84 cDNA and  
 RT chromosome assignment of related sequences.";  
 RL Gene 56:29-40(1987).  
 RN [5] SEQUENCE OF 1-311 FROM N.A.  
 RP MEDLINE=90076956; PubMed=2591742;  
 RA Walter T., Drabent B., Krebs H., Tomalak M., Heiss S.,  
 RA Benneke B.J.;  
 RT "Cloning and analysis of a human 86-kDa heat-shock-protein-encoding  
 RT gene.";  
 RL Gene 83:105-115(1989).  
 RN [6] SEQUENCE OF 184-731 FROM N.A.  
 RP TISSUE=Placenta;  
 RC MEDLINE=22388257; PubMed=12477932;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Coults B., Shewter K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.U., Useth T.B., Toshyuk S., Carninci P., Prange C.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mallaty S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Hulik S.W.,  
 RA Villalón D.K., Murthy D.W., Sodegren R.J., Lu X., Gibbs R.A.,  
 RA Fahy J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,  
 RA Blakesley R.W., Touchman J.W., Green B.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalusz D.E.,  
 RA Scherch A., Schein J.E., Jones S.J.M., Mair M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length  
 RT human and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RN [7] SEQUENCE OF 538-731 FROM N.A.  
 RP TISSUE=Heart;  
 RA Tanaka M., Tanaka T., Mitsui Y., Yamamoto M., Wood J.N.;  
 RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.  
 RN [8] SEQUENCE OF 1-20, AND PHOSPHORYLATION.  
 RP MEDLINE=89123325; PubMed=2492519;  
 RA Lees-Miller S., Anderson C.W.;  
 RT "Two human 90-kDa heat shock proteins are phosphorylated in vivo at  
 RT conserved serines that are phosphorylated in vitro by casein kinase  
 RT I.";  
 RL J. Biol. Chem. 264:2431-2437(1989).  
 RN [9] PHOSPHORYLATION BY DS-DNA KINASE.  
 RP MEDLINE=90008887; PubMed=2507541;  
 RA Lees-Miller S., Anderson C.W.;  
 RT "The human double-stranded DNA-activated protein kinase phosphorylates  
 RT the 90-kDa heat-shock protein, hsp90 alpha at two NH2-terminal  
 RT threonine residues.";  
 RL J. Biol. Chem. 264:17275-17280(1989).  
 RN [10] INTERACTION WITH OM34.  
 RP MEDLINE=9832497; PubMed=9660753;  
 RA Young J.C., Obermann W.M., Hartl F.U.;  
 RT "Specific binding of tetratricopeptide repeat proteins to the

RT C-terminal 12-kDa domain of hsp90.";  
 RL J. Biol. Chem. 273:18007-18010(1998).  
 RN [11] X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF 10-222.  
 RP MEDLINE=97262065; PubMed=9108479;  
 RA Stebbins C.E., Russo A.A., Schneider C., Rosen N., Hartl F.U.,  
 RA Pavletich N.P.;  
 RT "Crystal structure of an Hsp90-geldanamycin complex: targeting of a  
 RT protein chaperone by an antitumor agent.";  
 RL Cell 89:239-250(1997).  
 RN [12] X-RAY CRYSTALLOGRAPHY (1.5 ANGSTROMS) OF 10-222.  
 RP MEDLINE=99034582; PubMed=9817749;  
 RA Obermann W.M., Sondermann H., Russo A.A., Pavletich N.P., Hartl F.U.;  
 RT "In vivo function of Hsp90 is dependent on ATP binding and ATP  
 RT hydrolysis.";  
 RL J. Cell Biol. 143:901-910(1998).  
 CC -1- FUNCTION: Molecular chaperone. Has ATPase activity  
 CC (By similarity).  
 CC -1- SUBUNIT: Homodimer. Interacts with OM34.  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.  
 CC -1- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 90 FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL, X15183; CAA13259.1; -;  
 DR EMBL, X07270; CAA30255.1; -;  
 DR EMBL, M27024; AAA63194.1; -;  
 DR EMBL, M30626; AAA63023.1; -;  
 DR EMBL, BC000987; AAH00987.1; -;  
 DR EMBL, D87666; BAAL3430.1; -;  
 DR EMBL, D87666; BAAL3431.1; -;  
 DR PIR, A32319; HHHD86.  
 DR PDB, 1YER; 22-APR-98.  
 DR PDB, 1YES; 22-APR-98.  
 DR PDB, 1YET; 22-APR-98.  
 DR PDB, 1BYQ; 28-OCT-98.  
 DR TRANSFAC, T00992; -;  
 DR Genew; HGNC:5253; HSPCA.  
 DR MIM, 140571; -;  
 DR GO, GO:0005737; C:cytoplasm; NAS.  
 DR GO, GO:0003773; F:heat shock protein activity; NAS.  
 DR InterPro, IPR003594; AtPbind\_Atpase.  
 DR InterPro, IPR001404; Hsp90.  
 DR Pfam, PF02518; HATPase\_C\_1.  
 DR Pfam, PF00183; HSP90; 1.  
 DR PRINTS, PR00775; HEATSHOCK90.  
 DR SMART, SM00387; HATPase\_C; 1.  
 DR PROSITE, PS00298; HSP90; 1.  
 KM Chaperone, ATP-binding; Heat shock; Phosphorylation; 3D-structure.  
 FT INIT\_MET 0 0  
 FT MOD\_RES 4 4  
 FT MOD\_RES 6 6  
 FT MOD\_RES 230 230  
 FT MOD\_RES 262 262  
 FT MOD\_RES 262 262  
 FT CONFLICT 62 62  
 FT STRAND 17 20  
 FT STRAND 23 34  
 FT HELIX 40 41  
 FT HELIX 42 64  
 FT HELIX 46 69  
 FT HELIX 70 71  
 FT TURN 77 82  
 FT STRAND 83 86  
 FT STRAND 87 92  
 FT HELIX 99 103  
 FT TURN 104 107  
 T -> S (IN REF. 3, 4 AND 5).  
 PHOSPHORYLATION (BY DS-DNA KINASE).  
 PHOSPHORYLATION (BY DS-DNA KINASE).  
 PHOSPHORYLATION.  
 PHOSPHORYLATION.  
 T -> S (IN REF. 3, 4 AND 5).

```

FT HELIX 110 122
FT TURN 123 124
FT HELIX 127 133
FT TURN 134 134
FT HELIX 136 142
FT TURN 143 143
FT STRAND 144 152
FT TURN 154 155
FT STRAND 158 163
FT TURN 166 167
FT STRAND 168 173
FT STRAND 182 189
FT HELIX 191 197
FT TURN 199 209
FT TURN 211 212
FT STRAND 217 219
SQ SEQUENCE 731 AA; 84542 MW; 1249ABCFCE06297C CRC64;

Query Match 9.8%; Score 83.5; DB 1; Length 721;
Best Local Similarity 23.6%; Pred. No. 58;
Matches 39; Conservative 28; Mismatches 59; Indels 39; Gaps 7;

OY 6 FILNDTGEVSELKPHRTVTITQNGKMSSTVSEDFLLPYKGLKGYOPDGEWISG 65
DB 169 FTVRTDGE-----PMGRGKVIILHKEDQTEYLEERRI-----KEIVKHSQFIGYPTL 219
OY 66 F-EGGKDAQVYVNLSDPTFKPVFKKIEKKKEENKPTDVSKKQNPQVNHSQLNESHR 124
DB 220 FVEKERDK-----EVSDD-----EAEKEDKEKEKEKEKSEKSEDPET----- 257
OY 125 KEDLQREHSHQKSDSTKDTATVLDKN-----NLSKSTTNPN 163
DB 258 EDVGSDEBEERKDDKKKKKKIKKKYIDQBELNKTPIWTNP 301

RESULT 37
ARS2 DROME STANDARD; PRT; 943 AA.
AC Q9V9K7;
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Aresenite-resistance protein 2 homolog.
GN CG7843.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Gelinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter B.G., Helt G., Nelson C.R., Milos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Bencos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borovica D., Botchan M.R., Bouck J., Brockstein P., Brotner P.,
RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Doudon K., Doup L.B., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durkin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,

```

```

RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McInosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Paclet J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier B., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svitek R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Massaman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
RN [2]
RP REVISIONS, AND ALTERNATIVE SPLICING.
RC STRAIN=Berkley;
RX MEDLINE=22426069; PubMed=12537572;
RA Mista S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
RA Hradecky P., Huang Y., Kaminker J.S., Milburn G.H., Prochuk S.B.,
RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,
RA Belencourt B.R., Gelinker S.E., de Grey A.D.N.J., Drysdale R.A.,
RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.O.,
RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
RA Lewis S.E.;
RT "Annotation of the Drosophila melanogaster euchromatic genome: a
systematic review."
RL Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22(2002).
CC -1- ALTERNATIVE PRODUCTS:
CC Name=Long;
CC Name=Short;
CC Name=Q9V9K7-1; Sequence=Displayed;
CC IsoId=Q9V9K7-2; Sequence=VSP_000327;
CC -1- SIMILARITY: BELONGS TO THE ARS2 FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; AE003784; AAM68343.1; -
DR EMBL; AE003784; AAM68345.1; -
DR FlyBase; Fgn0033062; CG7843.
DR Pfam; PF04959; ARS2; 1.
KW Hypothetical protein; Alternative splicing.
FT VARSPLIC 47 50
FT Missing (in isoform short).
FT /FTId=VSP_000327.
SQ SEQUENCE 943 AA; 107221 MW; 0C1AF09B02E8AB08 CRC64;

Query Match 9.8%; Score 83.5; DB 1; Length 943;
Best Local Similarity 25.0%; Pred. No. 77;
Matches 36; Conservative 23; Mismatches 68; Indels 17; Gaps 4;

OY 14 EVSELKP-----HRTVTITQNGKMSSTV-----VSEDFLLPYKGLKGYOPDGW 61
DB 272 KYLDEKRPDPPVYERKABQMQSVKEKTKINSKEMSEADPVSTORKPVPVNSDGENW 331
OY 62 EISGPEGKDAQVYVNLSDPTFKPVFKKIEB-----KKEENKPTPVSKKQNPQVNHQ 118
DB 332 DDDDAENSA PKRLKEDSDSDSKPEDQKLNKKTKRRNRNSDDSSSSSSSSSBEK 391
OY 119 LNSHRKEDLQREHSHQKSDSTKD 142
DB 392 LREKYDVEDGLPAB--QKTEAERD 413

```

RESULT 38  
ID IF31 HUMAN STANDARD; PRT; 258 AA.  
AC 075822; 09802; Q9H802;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 15-SEP-2003 (Rel. 42, Last annotation update)  
DE Eukaryotic translation initiation factor 3 subunit 1 (eIF-3 alpha)  
DE (eIF3 p35) (eIF3j).  
GN EIF3J1.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
[1]  
RN SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
RX MEDLINE=99041954; PubMed=9822659;  
RA Block R.L., Vornlocher H.-P., Hershey J.W.B.;  
RT "Characterization of cDNAs encoding the p44 and p35 subunits of human  
RT translation initiation factor eIF3";  
RL J. Biol. Chem. 273:31901-31908(1998).  
[2]  
RN SEQUENCE FROM N.A.  
RC TISSUE=Ovarian carcinoma;  
RA Iwogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,  
RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,  
RA Magatsuma T., Hosokiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,  
RA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,  
RA Maranabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saio K.,  
RA Yamamoto J., Wakamatsu A., Nakamura Y., Nagahara K., Masuko Y.,  
RA Nimomiya K., Iwayanagi T.;  
RT "NEBO human cDNA sequencing project";  
RL Submitted (Aug-2000) to the EMBL/GenBank/DBJ databases.  
[3]  
RN SEQUENCE FROM N.A.  
RC TISSUE=uterus;  
RX MEDLINE=22388257; PubMed=12477932;  
RA Klausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Strasser R.D., Collins P.S., Wagner L., Shemen C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Scheffer C.F., Bhac N.K.,  
RA Hopkin R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh L.,  
RA Datchenko L., Marusina K., Farmer A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Donald M.F., Casavant T.L., Scheetz T.B.,  
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carrinci P., Prange C.,  
RA Rana S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,  
RA Boesek S.A., McGowan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahy J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
RA Buterfield Y.S.N., Krzywinski M.I., Skalska U., Smalls D.E.,  
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length  
RT human and mouse cDNA sequences";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
[4]  
RN FUNCTION: BINDS TO THE 40S RIBOSOME AND PROMOTES THE BINDING OF  
RN METHIONYL-TRNAI AND MRNA.  
[5]  
CC -1- SUBUNIT: EIF-3 IS COMPOSED OF AT LEAST 12 DIFFERENT SUBUNITS (by  
CC similarity).  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC EMBL, U97670; AAC78729.1; -  
CC EMBL, AK023388; BAB14555.1; -

DR EMBL: BC002719; AA02719.1; -  
DR Genew; HGNC:3270; EIF3J1.  
DR GK; 075822; -  
DR MIM; 603910; -  
DR GO; GO:0005852; C:eukaryotic translation initiation factor 3 . . . ; TAS.  
DR GO; GO:0003743; P:translation initiation factor activity; TAS.  
DR GO; GO:0006446; P:regulation of translational initiation; TAS.  
KW Initiation factor; Protein biosynthesis.  
FT DOMAIN 2 8 POLY-ALA.  
FT DOMAIN 29 32 POLY-GLY.  
FT DOMAIN 53 57 POLY-ASP.  
FT DOMAIN 218 224 POLY-LYS.  
FT CONFLICT 40 40 E -> G (IN REF. 1).  
FT CONFLICT 141 141 A -> T (IN REF. 2).  
SQ SEQUENCE 258 AA; 29062 MW; 8362423542445AA CRC64;  
  
Query Match 9.8%; Score 83; DB 1; Length 258;  
Best Local Similarity 25.4%; Pred. No. 21;  
Matches 30; Conservative 25; Mismatches 51; Indels 12; Gaps 4;  
  
QY 48 YNGLEKGYQFPQWISGREGKDAAY--VIMLSKOTFIKPYFKIIEKKKEENKPTFV 105  
DB 39 WEDEDEBDVKNMDDDEKKEKAEVKEVIXEK--KIAEKIKERROQKQKEI 95  
QY 106 SKKKNPQ-----VNSQLNESHREKDLQREHSQKSDTKDYATV--LDKNNISK 156  
DB 96 KRLLEPPEPKVLTPEEQALDKRLKQLQESDLEAKETPGVNNNAVIGIDANPESR 153  
  
RESULT 39  
ID TONB\_HAEIN STANDARD; PRT; 270 AA.  
AC P42872;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE TonB protein.  
GN TONB OR H10251.  
OS Haemophilus influenzae.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;  
OC Pasteurellaceae; Haemophilus.  
OX NCBI\_TaxID=727;  
[1]  
RN SEQUENCE FROM N.A.  
RC STRAIN=NTM1 TN106;  
RX MEDLINE=94245357; PubMed=8186372;  
RA Jarosik G.P., Sanders J.D., Cope L.D., Muller-Eberhard U.,  
RA Hansen E.J.;  
RT "A functional tonB gene is required for both utilization of heme and  
RT virulence expression by Haemophilus influenzae type b.";  
RL Infect. Immun. 62:2470-2477(1994).  
[2]  
RN SEQUENCE FROM N.A.  
RC STRAIN=Rd / KM20 / ATCC 51907;  
RX MEDLINE=95350630; PubMed=7542800;  
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,  
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,  
RA McInerney K., Suton G., Fitzhugh W., Fields C.A., Gocayne J.D.,  
RA Scott J.D., Shilly C.A., Spriggs T., Hedblom E., Cotton M.D.,  
RA Weisman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,  
RA Ueberback T.R., Hanna M.C., Nguyen D.T., Saudek D.W., Brandon R.C.,  
RA Fine L.D., Fritchman J.L., Fuhmann J.L., Geoghegan N.S.W.,  
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,  
RA Venter J.C.;  
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae  
RT Rd";  
RT Science 269:496-512(1995).  
[3]  
CC -1- FUNCTION: INTERACTS WITH OUTER MEMBRANE RECEPTOR PROTEINS THAT  
CC CARRY OUT HIGH-AFFINITY BINDING AND ENERGY DEPENDENT UPTAKE INTO  
CC THE PERIPLASMIC SPACE OF SPECIFIC SUBSTRATES. IT COULD ACT TO  
CC TRANSDUCE ENERGY FROM THE CYTOPLASMIC MEMBRANE TO SPECIFIC ENERGY-  
CC REQUIRING PROCESSES IN THE OUTER MEMBRANE, RESULTING IN THE  
CC RELEASE INTO THE PERIPLASM OF LIGANDS BOUND BY THESE OUTER





GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using SW model

Run on: February 10, 2004, 10:48:44 ; Search time 8.08902 Seconds

(without alignments)  
1011.574 Million cell updates/sec

Title: US-10-067-385-8\_COPY\_600\_773

Perfect score: 897  
Sequence: 1 KIVKDFARNTTVEFILNK.....ATVLDKNISKSTNNPNK 174

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : SwissProt\_41:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	100	11.1	1875	1	MLP1_YEAST
2	99	11.0	1398	1	TOP2_PLARK
3	97.5	10.9	1345	1	YHO0_YEAST
4	97.5	10.9	2468	1	MAPB_HUMAN
5	96	10.7	893	1	GYRA_STAEP
6	95	10.6	348	1	CYL2_HUMAN
7	95	10.6	1202	1	DPOW_ASCIM
8	93.5	10.4	649	1	HS70_PYRSA
9	93	10.4	715	1	ADSV_HUMAN
10	93	10.4	2464	1	MADP_MOUSE
11	92.5	10.3	443	1	DACA_BACSU
12	92.5	10.3	1702	1	IGAZ_HAETN
13	92.5	10.3	1888	1	YDT2_SCHPO
14	92	10.3	406	1	NSB1_MOUSE
15	92	10.3	1332	1	SPT7_YEAST
16	90	10.0	688	1	LIP_STAEP
17	90	10.0	778	1	YFK8_YEAST
18	89.5	10.0	433	1	TYG_STAEP
19	89.5	10.0	1066	1	PIKI_YEAST
20	89.5	10.0	1694	1	ICAO_HAETN
21	88	9.8	853	1	YCG1_YEAST
22	87.5	9.8	360	1	FENR_PEA
23	87.5	9.8	655	1	STO1_YEAST
24	87.5	9.8	678	1	GARP_PLAEP
25	87.5	9.8	879	1	RA50_PYRHO
26	87	9.7	6632	1	UN89_CAEEL
27	86.5	9.6	1006	1	RAV1_YEAST
28	86.5	9.6	1251	1	RBP2_PLAEP
29	86	9.6	200	1	SYST_LYCS
30	86	9.6	443	1	GLNA_PYRKO
31	86	9.6	914	1	PRPA_BACSU
32	86	9.6	5596	1	MDN1_HUMAN
33	85.5	9.5	752	1	DRS1_YEAST

34	85.5	9.5	1007	1	RGAL_YEAST	P39083 saccharomyc
35	85.5	9.5	2459	1	MAPB_RAT	P15205 ratuys norv
36	85	9.5	365	1	FENR_MESECR	P41343 mesembryant
37	85	9.5	472	1	6RED_LACLA	O9CNU6 lactococcus
38	85	9.5	657	1	Y040_MYCPN	P75062 mycoplasma
39	85	9.5	719	1	YBLE_SCHPO	Q10342 schizosacch
40	85	9.5	720	1	IF2_STAEP	O8CET4 staphylococ
41	85	9.5	895	1	RA50_THRYO	P58302 thermoplas
42	84.5	9.4	950	1	IF2_LACLC	O9X764 lactococcus
43	84	9.4	510	1	HGV2_HALRO	Q02508 halocynthia
44	84	9.4	796	1	FEC3_YEAST	P43573 saccharomyc
45	83.5	9.3	363	1	FENR_VICFA	P41346 vicia faba

## ALIGNMENTS

RESULT 1  
MLP1\_YEAST STANDARD; PRT; 1875 AA.  
ID AC 002455;  
DT 01-0CT-1993 (Rel. 27, Created)  
DT 01-0UN-1994 (Rel. 29, Last sequence update)  
DT 16-0CT-2001 (Rel. 40, Last annotation update)  
DR Myosin-like protein MLP1.  
GN MLP1 OR YKR095W OR YKR415.  
OS Saccharomyces cerevisiae (baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
OX NCBI\_TaxID=4932;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=S288c;  
RX MEDLINE=93247549; PubMed=8483450;  
RA Koelling R., Nguyen T., Chen E.Y., Botstein D.;  
RT "A new yeast gene with a myosin-like heptad repeat structure.";  
RL Mol. Gen. Genet. 237:359-369 (1993).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=94205265; PubMed=8154186;  
RA Bou G., Esteban P.F., Baladron V., Gonzalez G.A., Cantalejo J.G.,  
RA Remacha M., Jimenez A., del Rey P., Ballesta J.P.G., Revuelta J.L.;  
RT "The complete sequence of a 15,820 bp segment of Saccharomyces  
cerevisiae chromosome XI contains the UBI2 and MLP1 genes and three  
new open reading frames.";  
RL Yeast 9:1349-1354 (1993).  
CC - FUNCTION: MYOSIN-LIKE PROTEIN THAT IS PROBABLY INVOLVED IN DNA  
REPAIR.  
CC - SIMILARITY: SOME, TO THE TPR ONCOGENE.  
CC - CAUTION: REF.2 MISQUOTES THE GENE NAME AS "MPL1".  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
EMBL: L01992; AAA34783.1; -;  
EMBL: X73541; CAAS1948.1; -;  
EMBL: Z28320; CA82174.1; -;  
DR PIR: S38173; S38173.  
DR SGD: S0001803; MLP1.  
DR GO: GO:0005635; C:nuclear membrane; IDA.  
DR GO: GO:0005654; C:nucleoplasm; IDA.  
DR GO: GO:0006006; P:protein-nucleus import; IDA.  
KW Coiled coil; DNA repair.  
FT DOMAIN 69 487 COILED COIL (POTENTIAL).  
FT DOMAIN 531 1678 COILED COIL (POTENTIAL).  
FT DOMAIN 1834 1866 COILED COIL (POTENTIAL).  
FT CONFLICT 301 301 R -> A (IN REF. 1).  
SO SEQUENCE 1875 AA; 218455 MW; 683A0D34C906667 CRC64;

	Query Match	11.1%	Score 100,	DB 1,	Length 1875;
	Best Local Similarity	24.4%;	Pred. No. 14;		
	Matches	Conservative 38;	Mismatches 71;	Indels 46;	Gaps 10;
OY	4 VKDFARNTTVAFKFLNKDTGSEVSELPKPRVTYTIQNGEMSSSTIYSEEDF-----ILPVY 58	:	:	:   :	:
Dd	647 ISQITRSTETMSLTK---EIQDIYDKSKDPSITLGKESKRLLAEERFLSLTDLDT 703	:	:	:   :	:
OY	59 KGE--LEKGQFDSCMEISGFEGKDA-----GYINLSKDTFIKVPFKIKBER---- 104	:	:	:	:
Dd	704 KAENDQLAKRPDY-----LQNTLIKDSKTHTLETNYEVCSCKSLIVETHELINTKEEQKLK 759	:	:	:	:
OY	105 -----KEENKKPTFDVSKKKDNQVVAHSLNSHR-XEDP-----OREBHSQKSDTKOV 153	:	:	:	:
Dd	760 VHLKENLKQDELNK-----ISPERSLRINWTOQLQTKREDLLBEETRKSQQKXIDLEDA 815	:	:	:	:
OY	154 TA-----TVLDKNNISKSSTTNPNP 173	:	:	:	:
Dd	816 LSELKETTSQKDHHIKOLEEDNNNSN 840	:	:	:	:

RESULT 2			
TOP2_PLAFLK	STANDARD;	PRT;	1398 AA.
ID	TOP2_PLAFLK		
AC	P41001;		
DT	01-FEB-1995 (Rel. 31, Created)		
DT	01-FEB-1995 (Rel. 31, Last sequence update)		
DT	30-MAY-2000 (Rel. 39, Last annotation update)		
DE	DNA topoisomerase II (BC 5.99.1.3).		
GN	TOP2.		
OS	Plasmodium falciparum (isolate KI / Thailand).		
OC	Eukaryota, Alveolata, Apicomplexa, Haemosporidia; Plasmodium.		
OX	NCBI_TaxID=5839;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=94316496; PubMed=8041616;		
RA	Cheesman S., McAliese S., Goman M., Johnson D., Horrocks P.,		
RA	Ridley R.G., Kilbey B.J.;		
RT	"The gene encoding topoisomerase II from Plasmodium falciparum.";		
RL	Nucleic Acids Res. 22:2547-2551(1994).		
CC	-1- FUNCTION: CONTROL OF TOPOLOGICAL STATES OF DNA BY TRANSIENT		
CC	BREAKAGE AND SUBSEQUENT REJOINING OF DNA STRANDS. TOPOISOMERASE II		
CC	MAKES DOUBLE-STRAND BREAKS.		
CC	-1- CATALYTIC ACTIVITY: ATP-dependent breakage, passage and rejoining		
CC	of double-stranded DNA.		
CC	-1- SUBUNIT: Homodimer (By similarity).		
CC	-1- SUBCELLULAR LOCATION: Nuclear.		
CC	-1- MISCELLANEOUS: EUKARYOTIC TOPOISOMERASE I AND II CAN RELAX BOTH		
CC	NEGATIVE AND POSITIVE SUPERCOILS, WHEREAS PROKARYOTIC ENZYMES		
CC	RELAX ONLY NEGATIVE SUPERCOILS.		
CC	-1- SIMILARITY: Belongs to the type II topoisomerase family.		
CC	-----		
CC	CC This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (See <a href="http://www.ebi.ac.uk/announcements">http://www.ebi.ac.uk/announcements</a>		
CC	or send an email to <a href="mailto:license@ebi.ac.uk">license@ebi.ac.uk</a> ).		
CC	-----		
DR	EMBL; X79345; -; NOT_ANNOTATED_CDS.		
DR	HSSP; P06786; 1BGW.		
DR	InterPro; IPR003594; ATPbind_ATPase.		
DR	InterPro; IPR003957; CBFA_NFYA_topis.		
DR	InterPro; IPR001241; DNA_topoisolt.		
DR	InterPro; IPR002205; DNA_topoisolv.		
DR	Pfam; PF00204; DNA_gyraseb.1.		
DR	Pfam; PF00521; DNA_topoisolv.1.		
DR	Pfam; PF02518; HATase_C.1.		
DR	PRINTS; PR00615; CCAATSUBUNTA.		
DR	PRINTS; PR00418; TP12FAMILY.		
DR	ProDom; PD000742; DNA_topoisolv.1		
DR	ProDom; PD000742; TP12FAMILY.		

DR	SMART; SM00387; HATPase_C; 1.
DR	SMART; SM00433; TOP2C; 1.
DR	SMART; SM00434; TOP4C; 1.
DR	PROSITE; PS00177; TOPOISOMERASE_II; 1.
KW	Isomerase; Topoisomerase; DNA-binding; ATP-binding; Nuclear protein.
FT	NP BIND 144 149 ATP (POTENTIAL).
FT	ACT SITE 830 830 DNA CLEAVAGE (BY SIMILARITY).
FT	DOMAIN 271 281 POLY-ASN.
FT	DOMAIN 308 316 POLY-ASN.
FT	DOMAIN 1089 1093 POLY-LYS.
FT	DOMAIN 1227 1234 POLY-LYS.
SQ	SEQUENCE 1398 AA; 161029 MW; BAAD7BEE88FE5BE9 CRC64;

```

Query March 11.04; Score 99; DB 1; Length 1398;
Best Local Similarity 22.28; Pred. No. 12;
Matches 44; Conservative 45; Mismatches 67; Indels 42; Gaps 9;

QY 1 KIVVDFR-----NTVKEFLINKDGEVSELKPRVATVTLQNGKMSSTIVSEB 51
DB 1093 KVALVELYKGYDPYDKINIKIKKEBIFBQBLDAD-NPF-----DDEELIAGITVNDY 1145
QY 52 DFIL--PYVKGELKGYQFDGMEISFGKKGADGAVINLSKDTFIKVPFKXIE----- 103
DB 1146 DYLSMIPFSLITLK--VEDLTQLKEKRELEILANITVETWMLMDIENVEBALERQ 1202
QY 104 -----KKEENKPTFDVSKKKDNQVNHSQLNESHRKEDIQREHSQ--KSDSTKDVTA 155
DB 1203 NVELSNRESNK--FKVARQ-----GPSSMKKKKKKKKXSSDESECGDTSDSSEFLVN 1255
QY 156 TVLDKNNISSKSTTNPN 173
DB 1256 TLNKKVTKTKTTSSNN 1273

RESULT 3
YH00_YEAST STANDARD; PRT; 1345 AA.
AC P38800;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE Hypothetical 149.7 kDa protein in IRE1-KSP1 intergenic region.
YH080C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN 11
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RX MEDLINE=94378003; Pubmed=8091229;
RA Johnson M., Andrews S., Brinkman R., Cooper J., Ding H., Dover J.,
RA Du Z., Favellio A., Fulton L., Gelling S., Giesel C., Kirsten J.,
RA Kucaba T., Hillier L., Jier M., Johnston L., Langston Y.,
RA Lareille P., Louis E.J., Macri C., Mardis E., Meneses S., Mouser L.,
RA Nman M., Rifkin L., Riles J., St Peter H., Trevaakis B., Vaughan X.,
RA Vignati D., Wilcox L., Wohlman P., Waterston R., Wilson R.,
RA Vaudin M.;
RT "Complete nucleotide sequence of Saccharomyces cerevisiae chromosome
RT VIII."
RL Science 265:2077-2082(1994).
CC -1- SIMILARITY: TO YEAST YFL042C.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.1sb-sib.ch/announce/
CC or send an email to license@1sb-sib.ch).
CC
CC EMBL; U10556; AAB6895.1; -
CC
CC PIR; S46817; S46817.
CC

```



DR SGD; 50001122; YHR080C.  
DR InterPro; IPR004182; GRAM\_dom.  
DR Pfam; PF02893; GRAM; 1.  
DR SMART; SM00568; GRAM; 1.  
DR Hypothetical protein; Transmembrane.  
FT TRANSMEM 1198 1218 POTENTIAL.  
SQ SEQUENCE 1345 AA; 149679 MW; 2FDAB94A686564C2 CRC64;  
Query March 10.9%; Score 97.5; DB 1; Length 1345;  
Best Local Similarity 27.0%; Pred. No. 15; Mismatches 36; Indels 33; Gaps 6;  
Matches 33; Conservative 18; Mismatches 36; Indels 33; Gaps 6;  
QY 59 KGBLEKGYQPDGWEISGFEGRK-DAGYVINKSDTFIKPVKLEKKEENKPTFDVSK 117  
DB 1109 KGATEKG-----SVEGKQVSDVWLSRLDI-----SKAKSKKPVKQKMK 1149  
QY 118 KKDNPQVNSQLNSHREKDLQREHSQKSDSTQDVTATLD--KNNISKSTN--NP 172  
DB 1150 SHDGRPHFSKVE-----QKSESRKSDNDNDILTHIDPQNNSSSEIFNNKLLSP 1201  
QY 173 NK 174  
DB 1202 QK 1203  
RESULT 4  
MAPB\_HUMAN STANDARD; PRT; 2468 AA.  
AC P46821;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Microtubule-associated protein 1B (MAP 1B) [Contains: MAP1 light chain LC1].  
GN MAP1B.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=fetal brain;  
RX MEDLINE=95104835; PubMed=7806212;  
RA Lien L.L., Feener C., Fischbach N., Kunkel L.M.;  
RT "Cloning of human microtubule-associated protein 1B and the identification of a related gene on chromosome 15.";  
RL Genomics 22:273-280(1994).  
CC -1- FUNCTION: THE FUNCTION OF BRAIN MAPS IS ESSENTIALLY UNKNOWN.  
CC PHOSPHORYLATED MAP1B MAY PLAY A ROLE IN THE CYTOSKELETAL CHANGES  
CC THAT ACCOMPANY NEURITE EXTENSION. POSSIBLY MAP1B BINDS TO AT LEAST  
CC TWO TUBULIN SUBUNITS IN THE POLYMER, AND THIS BRIDGING OF SUBUNITS  
CC MIGHT BE INVOLVED IN NUCLEATING MICROTUBULE POLYMERIZATION AND IN  
CC STABILIZING MICROTUBULES.  
CC -1- SUBUNIT: 3 DIFFERENT LIGHT CHAINS, LC1, LC2 AND LC3, CAN ASSOCIATE  
CC WITH MAP1A AND MAP1B PROTEINS.  
CC -1- DOMAIN: Has a highly basic region with many copies of the sequence  
CC KKEE and KKEE/V, repeated but not at fixed intervals, which is  
CC responsible for the binding of MAP1B to microtubules.  
CC -1- PTM: LC1 is coexpressed with MAP1B. It is a polypeptide generated  
CC from MAP1B by proteolytic processing. It is free to associate with  
CC both MAP1A and MAP1B. It interacts with the amino-terminal region  
CC of MAP1B (By similarity).  
CC -1- SIMILARITY: TO MAP1A.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC -----  
CC EMBL; L06237; AAA18904.1; -.

DR Genew; HGNC:6836; MAP1B.  
DR MIM; 157129; -.  
DR GO; GO:0005875; C:microtubule associated complex; TAS.  
DR InterPro; IPR000102; MAP1B\_neuraxin.  
DR Pfam; PF00414; MAP1B\_neuraxin; 10.  
DR PROSITE; PS00230; MAP1B\_NEURAXIN; 6.  
KW Microtubules; Repeat; Phosphorylation.  
FT CHAIN ? 2468  
FT REPEAT 1878 1894 MAP1B 1.  
FT REPEAT 1895 1911 MAP1B 2.  
FT REPEAT 1912 1928 MAP1B 3.  
FT REPEAT 1929 1945 MAP1B 4.  
FT REPEAT 1946 1962 MAP1B 5.  
FT REPEAT 1963 1979 MAP1B 6.  
FT REPEAT 1997 2013 MAP1B 7.  
FT REPEAT 2014 2030 MAP1B 8.  
FT REPEAT 2031 2047 MAP1B 9.  
FT REPEAT 2048 2064 MAP1B 10.  
FT DOMAIN 589 790  
SQ SEQUENCE 2468 AA; 270618 MW; 540839C8DF09D461 CRC64;  
Query Match 10.9%; Score 97.5; DB 1; Length 2468;  
Best Local Similarity 24.8%; Pred. No. 29; Mismatches 47; Indels 33; Gaps 6;  
Matches 36; Conservative 29; Mismatches 47; Indels 33; Gaps 6;  
QY 14 KEFLNKDTGEVSLKPHRTVTITQNGKMSSTIVSEBDFILPVYKGELEKGYQPDGWEI 73  
DB 584 EKVMWKDKPVMTETKPSVTEKEVPSKEPS-----PV-KAEVA-----EK 623  
QY 74 SGFSGKQDAGYVINKSDTFIKPVKLEKKEENKPTFDVSKKDDNPQVNSQLNSH 133  
DB 624 QATDVKPPRAKEKTKVKEKTKVP-----EDKKEKPKGKGVAKKEDKTPI---KKEKP 675  
QY 134 RKEDLGR-----EHSQKSDSTKDY 153  
DB 676 KKEVKEKKEKKEKKEKKEKKEV 700  
RESULT 5  
GYRA\_STAEP STANDARD; PRT; 893 AA.  
AC P54112;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE DNA gyrase subunit A (EC 5.99.1.3).  
GN GYRA OR SE0005.  
OS Staphylococcus epidermidis.  
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.  
OX NCBI\_TaxID=1282;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 12228;  
RA Zhang Y., Ren S., Li H., Fu G., Lu L., Lu G., Jia J., Tu Y., Qin Z.,  
RA Chen Z., Wen Y.;  
RL Submitted (NOV-2002) to the EMBL/Genbank/DBJ databases.  
RN [2]  
CC SEQUENCE OF 1-94 FROM N.A., AND MUTAGENESIS OF SER-84.  
CC MEDLINE=92102204; PubMed=166207;  
CC Streedharan S., Peterson L.R., Fisher L.M.;  
CC "Ciprofloxacin resistance in coagulase-positive and -negative  
CC staphylococci: role of mutations at serine 84 in the DNA gyrase A  
CC protein of Staphylococcus aureus and Staphylococcus epidermidis.";  
CC Antimicrob. Agents Chemother. 35:2151-2154(1991).  
CC -1- FUNCTION: DNA GYRASE NEGATIVELY SUPERCOILS CLOSED CIRCULAR DOUBLE-  
CC STRANDED DNA IN AN ATP-DEPENDENT MANNER AND ALSO CATALYZES THE  
CC INTERCONVERSION OF OTHER TOPOLOGICAL ISOMERS OF DOUBLE-STRANDED  
CC DNA RINGS INCLUDING CATENANES AND KNOTTED RINGS.  
CC -1- CATALYTIC ACTIVITY: ATP-dependent breakage, passage and rejoining  
CC of double-stranded DNA.  
CC -1- SUBUNIT: MADE UP OF TWO CHAINS. THE A CHAIN IS RESPONSIBLE FOR DNA  
CC BREAKAGE AND REJOINING; THE B CHAIN CATALYZES ATP HYDROLYSIS. THE

CC ENZYME FORMS AN A2B2 TETRAMER.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation-

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

CC DR EMBL; AE016744; AA003602.1; -

CC DR EMBL; S72603; AA00672.1; -

CC DR PIR; A49832; A49832.

CC DR HSSP; P09097; IAB4.

CC DR InterPro; IPR002205; DNA\_topoisomIV.

CC DR Pfam; PF03989; DNA\_gyraseA\_C1\_6.

CC DR ProDom; PD000742; DNA\_topoisomIV\_1.

CC DR SMART; SM00434; TOP4c; 1.

CC DR TIGRFAMs; TIGR01063; gyra; 1.

CC KM Topoisomerase; Isomerase; DNA-binding; Antibiotic resistance;

CC KW Complete proteome.

CC FT ACT SITE 123 DNA CLEAVAGE (BY SIMILARITY).

CC FT MUTAGEN 84 S->P: RESISTANT TO CIPROFLOXACIN.

CC SQ SEQUENCE 893 AA; 100113 MW; 2A6A7CD345A526CE CRC64;

CC -----

CC Query Match 10.7%; Score 96; DB 1; Length 893;

CC Best Local Similarity 24.3%; Pred. No. 12;

CC Matches 43; Conservative 33; Mismatches 77; Indels 24; Gaps 8;

CC -----

CC QY 7 FARNTTVKPEFLNKGVESELKPHRYVTYIIONGKEMS-STIYSEDFILIPYKGELEKRG 65

CC DB 723 YKRTIPVSGYRSLNKGKGIK-----TATITERNGNIVCTTYGEEDLAVTNAGVI--- 775

CC QY 66 VQPDGMEISGFEKGDAGY-VINLSKDTPIKPYFKIEKKEEEN-----KETPDV 115

CC DB 776 IRLDVHDIISQ-NGRAAQGVRLMKLGDQGVSTVAKVNEEDDMEVNDAAQSTTTETADV 834

CC QY 116 SKKKNPOVNSQLNESHKEDLQREE-HSQKSDSTKYVTATVLDKNNTSSKSTNN 171

CC DB 835 EEVVD-----DQTPGNAIHTEGDAMESVSPENDRIDIRDFMDRVNEDISASDN 887

CC -----

CC RESULT 6

CC CYL2\_HUMAN STANDARD; PRT; 348 AA.

CC ID CYL2\_HUMAN

CC AC 014053;

CC DT 01-NOV-1997 (Rel. 35, Created)

CC DT 01-NOV-1997 (Rel. 35, Last sequence update)

CC DT 28-FEB-2003 (Rel. 41, Last annotation update)

CC DE Cytlicin II (Multiple-band polypeptide II).

CC OS CYL2 OR CYL2.

CC OC Homo sapiens (Human).

CC OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

CC OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

CC OX NCBI\_Taxid=9606;

CC RN [1]

CC RP SEQUENCE FROM N.A.

CC RC TISSUE=Testis;

CC RX MEDLINE=95255491; PubMed=7737358;

CC RA Hees H., Heid H., Zimbelmann R., Franke N.W.;

CC RT "The protein complexity of the cytoskeleton of bovine and human sperm

CC heads: the identification and characterization of cylicin II.";

CC RL Exp. Cell Res. 218:174-182(1995).

CC CC -1- FUNCTION: POSSIBLE ARCHITECTURAL ROLE DURING SPERMATOGENESIS. MAY

CC BE INVOLVED IN SPERMATID DIFFERENTIATION.

CC CC -1- SUBCELLULAR LOCATION: CALYX; SPERM HEAD CYTOSKELETAL STRUCTURE.

CC CC -1- TISSUE SPECIFICITY: Testis.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation-

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

CC DR EMBL; Z46788; CAA86752.1; -

CC DR PIR; I37271; I37271.

CC DR Genew; HGNC; 2583; CYL2.

CC DR MIM; 604035; -

CC DR GO; 0005200; F-structural protein; Repeat; Sperm Spermatogenesis.

CC KM Cytoskeleton; Structural protein; Repeat; Sperm Spermatogenesis.

CC FT DOMAIN 25 347 31 X 3 AA REPEATS OF K-K-X.

CC FT DOMAIN 157 240 3 X APPROXIMATE TANDEM REPEATS.

CC FT REPEAT 157 184 1.

CC FT REPEAT 185 212 2.

CC FT REPEAT 213 240 3.

CC SQ SEQUENCE 348 AA; 39079 MW; D86766599C1809E7 CRC64;

CC -----

CC Query Match 10.6%; Score 95; DB 1; Length 348;

CC Best Local Similarity 30.9%; Pred. No. 5.3;

CC Matches 38; Conservative 20; Mismatches 39; Indels 26; Gaps 7;

CC -----

CC QY 59 KGELEKGYQPDGMEISGFEKGDAGYVIMLSKDTPIKPYFKIEKKEEENKPTF---DV 115

CC DB 205 ESEGEKG---GTEDSKKGGKDS---KKGDSALHLDQVADKEDKDGKDAKGD 256

CC QY 116 SK-KKDNPOVNSQLN-----ESHKEDLQREEHSQKSDSTKD---VTATVLDKNNT 163

CC DB 257 SKDAKDAKEIKKKKKKKKPSSTSDSDVDYKE---SKKATKAKKAKKDTKEESA 313

CC QY 164 SSK 166

CC DB 314 DSK 316

CC -----

CC RESULT 7

CC DPOM\_ASCIM STANDARD; PRT; 1202 AA.

CC ID DPOM\_ASCIM

CC AC P2374;

CC DT 01-AUG-1991 (Rel. 19, Created)

CC DT 01-AUG-1991 (Rel. 19, Last sequence update)

CC DT 15-DEC-1998 (Rel. 37, Last annotation update)

CC DE Probable DNA polymerase (EC 2.7.7.7).

CC OS Ascobolus immerus.

CC OG Mitochondrion.

CC OC Plasmid pA12.

CC OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Pezizomycetes;

CC OX Pezizales; Ascobolaceae; Ascobolus.

CC OX NCBI\_Taxid=5191;

CC RN [1]

CC RP SEQUENCE FROM N.A.

CC RC STRAIN=2/1;

CC RX MEDLINE=90066356; PubMed=2573821;

CC RA Kempken F., Weinhardt F., Esser K.;

CC RT "In organello replication and viral affinity of linear,

CC extrachromosomal DNA of the ascomycete Ascobolus immerus.";

CC RL Mol. Gen. Genet. 218:523-530(1989).

CC CC -1- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate = N diphosphate

CC + (DNA) (N).

CC CC -1- MISCELLANEOUS: THIS DNA POLYMERASE REQUIRES A PROTEIN AS A PRIMER

CC (BY SIMILARITY).

CC CC -1- SIMILARITY: BELONGS TO DNA POLYMERASE TYPE-B FAMILY. STRONG, TO

CC DNA POLYMERASE OF OTHER FUNGAL AND PLANTS MITOCHONDRIAL PLASMIDS.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation-

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

CC DR EMBL; X15982; CAA34106.1; -

CC DR PIR; S05362; S05362.

CC DR InterPro; IPR006172; DNA\_pol\_B.

DR InterPro: IPR004868, DNA\_pol\_B\_2.  
 DR Pfam: PF03175, DNA\_pol\_B\_2; 1.  
 DR SMART: SM00486; POLB; 1.  
 DR PROSITE: PS00116; DNA POLYMERASE B; 1.  
 KM Transferrase; DNA-directed DNA polymerase; DNA replication;  
 KM DNA-binding; Plasmid; Mitochondrion.  
 SQ SEQUENCE 1202 AA; 138279 MW; 51D41FCBDBF2CDE CRC64;

Query Match 10.6%; Score 95; DB 1; Length 1202;  
 Best Local Similarity 22.8%; Pred. No. 20;  
 Matches 44; Conservative 37; Mismatches 80; Indels 32; Gaps 9;

QY 10 NTTVVEFLINK---DTGEVSELKPHRYVTYVTONGKMSSTI---VSEEDF--ILPYK 60  
 DB 305 NTFQPFVNNAKIKPTTGNVNSIGFQNTV-TLTDVETLIKTLAIFLEREDHTVMSYDEG 363  
 QY 61 ELEK-----GYQFGWEISGFEQKQAGYVNLKDPFIKPVFKIE-----EKKEE 108  
 DB 364 DIDEKFPKSGISFPPKPKTEGKTKYANTTPPIKDIYVQINKLNNGDLKRTMDL 423  
 QY 109 NK-PTFVSKKKQNPQVNSQLNSHRKEDLQREHSQKSDSTKDVATVLDKNNI---- 163  
 DB 424 SKWPLKLNKQKTSGEIRMTIKNNKNGSYDI--IGHMIINDENVTTRAVDNSIKIF 481  
 QY 164 ---SSKSTNNPN 173  
 DB 482 TVTDSMGNTNDN 494

RESULT 8  
 HS70\_PYRSA STANDARD; PRT; 649 AA.

AC P37859;  
 DT 01-OCT-1994 (Rel. 30, Created)  
 DT 01-OCT-1994 (Rel. 30, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Heat shock 70 kDa protein.  
 GN HSP70.  
 OS Pyrenomonas salina.  
 OC Nucleomorph.  
 OC Eukaryota; Cryptophyta; Cryptomonadaceae; Pyrenomonas.  
 OX NCBI\_TaxID=3034;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MBL0196=94268506; PubMed=8208251;  
 RA Hofmann C.J.B., Rensing S.A., Haebler M.M., Martin W.F., Mueller S.B.,  
 RA Couch J., McFadden G.I., Igloi G.L., Walter U.-G.,  
 RT "The smallest known eukaryotic genomes encode a protein gene: towards  
 an understanding of nucleomorph functions.";  
 RL Mol. Gen. Genet. 243:600-604(1994).  
 CC -1- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 70 FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: X72621; CA51197.1; .  
 DR PIR: S42488; S42488.  
 DR HSP; P08109; ICRK.  
 DR InterPro: IPR001023; Hsp70.  
 DR Pfam: PF00012; HSP70; 1.  
 DR PRINTS: PRO0301; HEATSHOCK70.  
 DR PRODOM: PD000089; HSP70; 1.  
 DR PROSITE: PS00297; HSP70\_1; 1.  
 DR PROSITE: PS00329; HSP70\_2; 1.  
 DR PROSITE: PS01036; HSP70\_3; 1.  
 KM ATP-binding; Heat shock; Nucleomorph.  
 SQ SEQUENCE 649 AA; 72079 MW; B627B08FP90C9164 CRC64;

Query Match 10.4%; Score 93.5; DB 1; Length 649;  
 Best Local Similarity 24.5%; Pred. No. 13;  
 Matches 38; Conservative 26; Mismatches 50; Indels 41; Gaps 7;

QY 17 IINAKTGEVSELKPHRYVTYVTONG---KMSSTIYSEEDFILPYVGGELKQGFQGMKI 73  
 DB 491 IINVASDSKTSKSNKITTTNDKGRUSKEIRMEEAR---KYTFDEK----- 537  
 QY 74 SGEFGKQA-----GYVNLIS---KDTPIKPVFKIEEKKEENKPTFVSKKKQNPVN 125  
 DB 538 --LDRKLEAKNSLENYAVNIRVTD-----EKIKKEIQEEDKKISIEKVE----- 582  
 QY 126 HSQLNESHKEDLQREHSQKSDSTKDVATVATVLDK 160  
 DB 583 --VLSPRTNDELEKRYEKEKEKLNPNPNIISK 615

RESULT 9  
 AD5V\_HUMAN STANDARD; PRT; 715 AA.

ID AD5V\_HUMAN  
 AC Q9YED3; O8WU97;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Adseverin (Scinderin).  
 GN SCIN.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Skin;  
 RX MEDLINE=22388257; PubMed=12477932;  
 RA Strausberg R.L., Feingold B.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Scheffer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Uedin T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullen S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Villalón D.K., Muzny D.M., Sodergren B.J., Lu X., Gibbs R.A.,  
 RA Richardson S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hultyk S.W.,  
 RA Fahy J., Heltan E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green B.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schultz J., Myers R.M.,  
 RA Butlerfield Y.S.N., Krzywinski M.I., Skalska U., Smalins D.E.,  
 RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length  
 human and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RN [2]  
 RP SEQUENCE OF 1-527 FROM N.A.  
 RA Kalicki J., Smith-Craig R.;  
 RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE OF 248-715 FROM N.A.  
 RC TISSUE=Placenta;  
 RA Itoigai T., Oca T., Hayaishi K., Sugiyama T., Otsubi T., Suzuki Y.,  
 RA Nishikawa T., Nagai K., Sugano S., Takahashi-Fujii A., Hara H.,  
 RA Tanase T., Nomura Y., Toshiya S., Komai F., Hara R., Takeuchi K.,  
 RA Arita M., Nabekura T., Ishii S., Kawai Y., Saito K., Yamamoto J.,  
 RA Wakamatsu A., Nakamura Y., Nagahari K., Masuo Y., Oshima A.;  
 RT "NEBO human cDNA sequencing project.";  
 RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: Ca(2+)-dependent actin filament-severing protein that is  
 presumed to have a regulatory function in exocytosis by affecting  
 the organization of the microfilament network underneath the  
 plasma membrane. In vitro, also has barbed end capping and  
 nucleating activities in the presence of Ca(2+).

```

CC -!- SIMILARITY: BELONGS TO THE VILILIN/GELSOLIN FAMILY.
CC -!- SIMILARITY: Contains 6 gelsolin-like repeats.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: BC021090; AAH21090.1; -
CC EMBL: AC005281; AAD15423.1; -
CC EMBL: AK027778; BAB55361.1; -
CC HSSB: P02640; 2VIL.
CC InterPro: IPR001974; Gelsolin.
CC Pfam: PF00626; Gelsolin.6.
CC PRINTS: PR00597; GELSOLIN.
CC CYCOSKLEON; Actin-binding; Repeat; Calcium; Capping protein.
CC SMART; SMO0262; GEL; 6.
CC DOMAIN 1 363 ACTIN-SEVERING (POTENTIAL).
CC REPEAT 27 76 CA(2+)-DEPENDENT ACTIN BINDING.
CC REPEAT 148 188 GELSOLIN-LIKE 1.
CC REPEAT 265 307 GELSOLIN-LIKE 2.
CC REPEAT 398 451 GELSOLIN-LIKE 3.
CC REPEAT 523 564 GELSOLIN-LIKE 4.
CC REPEAT 626 668 GELSOLIN-LIKE 5.
CC SITE 112 119 POLYPHOSPHOINOSITIDE BINDING (BY
CC SIMILARITY).
CC SITE 138 146 POLYPHOSPHOINOSITIDE BINDING (BY
CC SIMILARITY).
CC CONFLICT 61 61 R -> H (IN REF. 2).
CC SEQUENCE 715 AA; 80508 MW; 45FBE42CBCHFDB0 CRC64;
CC
CC Query Match 10.4%; Score 93; DB 1; Length 715;
CC Best Local Similarity 21.0%; Pred. No. 16;
CC Matches 37; Conservative 39; Mismatches 58; Indels 42; Gaps 6;
CC
QY 3 VKKDFARNVTVEKPIFKKGVSELEKPHRTVTYTIQNGKMSSTIYSEEDFILPVYKGL 62
DB 239 IYADISNRKMAKLYMSDASGSM-----RVTVVAEENFPNMAMLSECFILID--HGA 290
QY 63 EKGYOFDGEISGFBEK-----KDAGVYINLSKDTFIK-----PVFKK----- 100
DB 291 KQIFWKKGDANPQERKAMKTAEBEPLQDMYTSKATQIQLVEGSGTPIEFKQPFQMDRK 350
QY 101 -----IEKKKEENKPTFDVSKKKDPQV--NHSQLESHRKEDLQREHS 144
DB 351 DQSDGFGKYVTEKVAQIKQIFPDASKLHSSPQMAQHNMVDDGSGKVEIMRVENN 406
CC
RESULT 10
MAPB_MOUSE STANDARD; PRT; 2464 AA.
AC P14873;
DT 01-APR-1990 (rel. 14, Created)
DT 01-APR-1990 (rel. 14, Last sequence update)
DT 28-FEB-2003 (rel. 41, Last annotation update)
DE Microtubule-associated protein 1B (MAP1B) (MAP1.2) (MAP1(X))
DE [Contains: MAP1 light chain LC1].
GN MAP1B OR MAP1B OR MAP1B OR MAP1B.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A., AND DOMAIN.
RC STRAIN=Swiss Webster; TISSUE=Brain;
RX MEDLINE=90094539; PubMed=2480963;
RA Noble M., Lewis S.A., Cowan N.J.;
RT "The microtubule binding domain of microtubule-associated protein
RT MAP1B contains a repeated sequence motif unrelated to that of MAP2

```

```

RT and tau."
CC J. Cell Biol. 109:3367-3376(1989).
CC -!- FUNCTION: THE FUNCTION OF BRAIN MAPS IS ESSENTIALLY UNKNOWN.
CC PHOSPHORYLATED MAP1B MAY PLAY A ROLE IN THE CYTOSKELETAL CHANGES
CC THAT ACCOMPANY NEURITE EXTENSION. POSSIBLY MAP1B BINDS TO AT LEAST
CC TWO TUBULIN SUBUNITS IN THE POLYMER, AND THIS BRIDGING OF SUBUNITS
CC MIGHT BE INVOLVED IN NUCLEATING MICROTUBULE POLYMERIZATION AND IN
CC STABILIZING MICROTUBULES.
CC -!- SUBUNIT: 3 DIFFERENT LIGHT CHAINS, LC1, LC2 AND LC3, CAN ASSOCIATE
CC WITH MAP1A AND MAP1B PROTEINS.
CC -!- DOMAIN: Has a highly basic region with many copies of the sequence
CC KKEE and KKEI/V, repeated but not at fixed intervals, which is
CC responsible for the binding of MAP1B to microtubules.
CC -!- PTM: LC1 IS COEXRESSED WITH MAP1B. IT IS A POLYPEPTIDE GENERATED
CC FROM MAP1B BY PROTEOLYTIC PROCESSING. IT IS FREE TO ASSOCIATE WITH
CC BOTH MAP1A AND MAP1B. IT INTERACTS WITH THE AMINO-TERMINAL REGION
CC OF MAP1B.
CC -!- SIMILARITY: TO MAP1A.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: X51396; CAA35761.1; -
CC PIR: S07549; QMSPL.
CC MGD; MG1:1306778; Mtap1b.
CC GO; GO:0016358; P: dendrite morphogenesis; IMP.
CC GO; GO:0001578; P: microtubule bundling; IMP.
CC InterPro: IPR000102; MAP1B_neuraxin.
CC Pfam; PF00414; MAP1B_neuraxin; 10.
CC PROSITE; PS00230; MAP1B_NEURAXIN; 7.
CC Microtubules; Repeat; Phosphorylation.
CC CHAIN 2 2464 MAP1 LIGHT CHAIN LC1.
CC REPEAT 1874 1890 MAP1B 1.
CC REPEAT 1891 1907 MAP1B 2.
CC REPEAT 1908 1924 MAP1B 3.
CC REPEAT 1925 1941 MAP1B 4.
CC REPEAT 1942 1958 MAP1B 5.
CC REPEAT 1959 1975 MAP1B 6.
CC REPEAT 1993 2009 MAP1B 7.
CC REPEAT 2010 2026 MAP1B 8.
CC REPEAT 2027 2043 MAP1B 9.
CC REPEAT 2044 2060 MAP1B 10.
CC DOMAIN 589 787 LYS-RICH (HIGHLY BASIC, CONTAINS MANY
CC KKEE AND KKEI/V REPEATS).
CC SEQUENCE 2464 AA; 270408 MW; FBD3DD99CFBDA87 CRC64;
CC
CC Query Match 10.4%; Score 93; DB 1; Length 2464;
CC Best Local Similarity 27.8%; Pred. No. 61;
CC Matches 49; Conservative 24; Mismatches 57; Indels 46; Gaps 11;
CC
QY 14 KEFILNKQGVSELEKPHRTVTYTIQNGKMSSTIYSEEDFILPVYKGELEKGYQDGEI 73
DB 584 EKVLYKDDPVYTESKP---SVT---EKVSS---KEEQ--SPV-KAEVA-----EK 623
QY 74 SGEFGKQAGVYINLSKDTFIKPVFK-KIEKKKEENKPTFDVSKKKDPQVNHSQLNES 132
DB 624 QATSEKP-----KTKQKVYKKEIKTKLEKKEE--KPKKEVKKEDTKPL--KDEK 672
QY 133 HKKEDLQRE-----EHSQKSDSTQVATYATLDKNISKSTTNPVK 174
DB 673 PKKEVYKKEIKKIEKKKEERKILKEVKKETPLKDAKKEVKEKKEKKEKPKK 728
CC
RESULT 11
DACA_BACSU STANDARD; PRT; 443 AA.
ID DACA_BACSU
AC P08750;
DT 01-AUG-1988 (rel. 08, Created)

```

DT 01-OCT-1994 (Rel. 30, last sequence update)  
 DT 28-FEB-2003 (Rel. 41, last annotation update)  
 DE D-alanyl-D-alanine carboxypeptidase precursor (EC 3.4.16.4) (DD-  
 DE peptidase) (DD-carboxypeptidase) (CPase) (PB55).  
 GN DCA.  
 OS Bacillus subtilis.  
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
 OK NCBI\_TaxID=1423;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=168;  
 RX MEDLINE=96051385; PubMed=7584024;  
 RA Ogasawara N., Nakai S., Yoshikawa H.;  
 RT "Systematic sequencing of the 180 kilobase region of the Bacillus  
 RT subtilis chromosome containing the replication origin.";  
 RL DNA Res. 1:1-14(1994).  
 RP [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=168;  
 RX MEDLINE=98044033; PubMed=9384377;  
 RA Kunze F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,  
 RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,  
 RA Borriss R., Boutsier L., Brans A., Braun M., Brignell S.C., Bron S.,  
 RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,  
 RA Choi S.K., Codani J.J., Conterton I.F., Cummings N.J., Daniel R.A.,  
 RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,  
 RA Enlian K.D., Errington J., Fadre C., Ferrari E., Foulger D.,  
 RA Fritz C., Fujita M., Fujita Y., Fuma S., Gallizzi A., Galleron N.,  
 RA Ghim S.Y., Glaser P., Goffeau A., Golightly B.J., Grandi G.,  
 RA Guisepi G., Guy B.J., Haga K., Hahsch J., Harwood C.R., Henaut A.,  
 RA Hilbert H., Holtsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,  
 RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,  
 RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kuman M.,  
 RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,  
 RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,  
 RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,  
 RA Noone D., O'Reilly M., Ogawa K., Ogihara A., Oudega B., Park S.H.,  
 RA Parro V., Pohl T.M., Portecelle D., Porwollik S., Prescott A.M.,  
 RA Priesack E., Pujic P., Purnelle B., Rapoport G., Ray M., Reynolds S.,  
 RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadate Y.,  
 RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,  
 RA Sekiguchi J., Sekowaka A., Seror S.J., Serro P., Shin B.S., Soldo B.,  
 RA Sorokin A., Tacconi B., Takagi T., Takahashi H., Takemaru K.,  
 RA Takouchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,  
 RA Toato V., Uchiyama S., Vandenbol M., Vannier P., Vassarotti A.,  
 RA Viari A., Wambut R., Wedler E., Wedler H., Weizenegger T.,  
 RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,  
 RA Yoshida K., Yoshikawa H.F., Zumbstein B., Yoshikawa H., Danchin A.;  
 RT "The complete genome sequence of the Gram-positive bacterium Bacillus  
 RT subtilis".  
 RL Nature 390:249-256(1997).  
 RN [3]  
 RP SEQUENCE OF 32-102.  
 RX MEDLINE=80182289; PubMed=6768745;  
 RA Waxman D.J., Strominger J.L.;  
 RT "Sequence of active site peptides from the penicillin-sensitive D-  
 RT alanine carboxypeptidase of Bacillus subtilis. Mechanism of  
 RT penicillin action and sequence homology to beta-lactamases.";  
 RL J. Biol. Chem. 255:3964-3976(1980).  
 RN [4]  
 RP SEQUENCE OF 103-443 FROM N.A.  
 RX MEDLINE=86250602; PubMed=3087956;  
 RA Todd J.A., Roberts A.N., Johnstone K., Pigot P.J., Winter G.,  
 RA Ellar D.J.;  
 RT "Reduced heat resistance of mutant spores after cloning and  
 RT mutagenesis of the Bacillus subtilis gene encoding penicillin-binding  
 RT protein 5.";  
 RL J. Bacteriol. 167:257-264(1986).  
 RN [5]  
 RP SEQUENCE OF 414-443.  
 RX MEDLINE=81117303; PubMed=6780559;  
 RA Waxman D.J., Strominger J.L.;  
 RT "Primary structure of the COOH-terminal membranous segment of a

RT penicillin-sensitive enzyme purified from two Bacilli.";  
 RL J. Biol. Chem. 256:2067-2077(1981).  
 CC -1- FUNCTION: REMOVES C-TERMINAL D-ALANYL RESIDUES FROM SUGAR-PEPTIDE  
 CC CELL WALL PRECURSORS.  
 CC -1- CATALYTIC ACTIVITY: D-alanyl-D-alanine + H(2)O = 2 D-alanine.  
 CC -1- PATHWAY: Peptidoglycan synthesis; final stages.  
 CC -1- SUBCELLULAR LOCATION: Membrane-associated.  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S11; ALSO KNOWN AS THE  
 CC D-ALANYL-D-ALANINE CARBOXYPEPTIDASE 1 FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; D26185; BAA05246.1; -  
 CC EMBL; 239104; CAB11786.1; -  
 CC EMBL; M33766; AAA22375.1; -  
 CC PIR; S66040; S66040.  
 CC MEROPS; S11.001; -  
 CC DR Subtilat; BG10074; dact.  
 CC DR InterPro: IPR001967; Ala/AlaCBPase1.  
 CC DR Pfam: PF00768; Peptidase\_S11; 1.  
 CC DR PRINTS; PR00725; DADACBPASB1.  
 CC DR HydroLase; Carboxypeptidase; Peptidoglycan synthetase; Cell wall;  
 CC Membrane; Signal; Complete proteome.  
 CC FT SIGNA1 1 31  
 CC FT CHAIN 32 443 D-ALANYL-D-ALANINE CARBOXYPEPTIDASE.  
 CC FT ACT SITE 67  
 CC FT CONFLICT 100 100 E -> Q (IN REF. 3).  
 CC FT CONFLICT 227 227 E -> Q (IN REF. 4).  
 CC SQ SEQUENCE 443 AA; 48636 MW; DA6C5B0307D7C117 CRC64;  
 CC  
 CC Query Match 10.3%; Score 92.5; DB 1; Length 443;  
 CC Best Local Similarity 25.0%; Pred. No. 10;  
 CC Matches 36; Conservative 26; Mismatches 49; Indels 33; Gaps 6;  
 CC  
 CC QY 6 DRA-RTTQKEPLINKDGEVSEIKPHRYVTITONGKESSTVSEDFILPYKGLK 64  
 CC DB 304 DYAIPNFMSKKEIYAECD---QVKGKH-TISVDKKEKEKRGVLTWKAASLPYKNG-EK 356  
 CC QY 65 GYQPGWMEISGREGKDDAVYINLSKDTFKIPVFKIEKKEBENKPTFQVSKKKNPOV 124  
 CC DB 357 NYKAK-----VTLNKNITLAPYKGTGK-----LTAEYTGDEK 391  
 CC QY 125 NNSQANSHKEDLQREHSQKSD 148  
 CC DB 392 DYGFANSLDLAGVDLYTKENVEKAN 415  
 CC  
 CC RESULT 12  
 CC IGA2\_HAEIN STANDARD; PRT; 1702 AA.  
 CC AC P45384;  
 CC DT 01-NOV-1995 (Rel. 32, Created)  
 CC DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 CC DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 CC DE Immunoglobulin A1 protease precursor (EC 3.4.21.72) (IGA1 protease).  
 CC GN IGA.  
 CC OS Haemophilus influenzae.  
 CC OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;  
 CC OC Pasteurellaceae; Haemophilus.  
 CC OK NCBI\_TaxID=727;  
 CC RN [1]  
 CC RP SEQUENCE FROM N.A.  
 CC RC STRAIN=HK715 / Serotype B;  
 CC RX MEDLINE=92234949; PubMed=1373717;  
 CC RA Poulsen K., Reinholdt J., Klilian M.;  
 RT "A comparative genetic study of serologically distinct Haemophilus  
 RT influenzae type 1 immunoglobulin A1 proteases.";

RL J. Bacteriol. 174:2913-2921 (1992).

CC -1- FUNCTION: VIRULENCE FACTOR; CLEAVES HOST IMMUNOGLOBULIN A

CC PRODUCING INTERACT PC AND FAB FRAGMENTS.

CC -1- CATALYTIC ACTIVITY: Cleavage of immunoglobulin A molecules at

CC certain Pro-I-Xaa bonds in the hinge region. No small molecule

CC substrates are known.

CC -1- SUBCELLULAR LOCATION: Secreted.

CC -1- DOMAIN: THE SIGNAL PEPTIDE GUIDES THE PRECURSOR TO THE PERIPLASMIC

CC SPACE, AND THE CARBOXY-TERMINAL HELPER DOMAIN ASSOCIATES WITH THE

CC OUTER MEMBRANE TO FORM A PORE FOR EXCRETION OF THE PROTEASE

CC DOMAIN. THE HELPER DOMAIN IS THEN RELEASED BY AUTOPROTEOLYSIS (BY

CC SIMILARITY).

CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S6.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL, M87489; AAA2966.1; -

DR PIR, A41859; A41859.

DR MEROPS, S06.001; -

DR InterPro, IPR006315; Autotransport.

DR InterPro, IPR005546; Autotransporter.

DR InterPro, IPR000710; IGA\_S6.

DR InterPro, IPR004899; Pertactin.

DR Pfam, PF03797; Autotransporter; 1.

DR Pfam, PF03295; IGAL; 1.

DR Pfam, PF03212; Pertactin; 1.

DR PRINTS, PR00921; IGASERPTASE.

DR TIGRFAMs, TIGR01414; autotrans\_bar1; 1.

KW Hydrolase; Serine protease; Transmembrane; Zymogen; Repeat; Signal.

FT SIGNAL 1 25 POTENTIAL.

FT CHAIN 26 1014 IMMUNOGLOBULIN A1 PROTEASE.

FT PROSER 1015 1702 HELPER PEPTIDE (POTENTIAL).

FT ACT\_SITE 288 288 PROBABLE.

FT DOMAIN 1109 1124 2 X 8 AA TANDEM REPEATS OF A-K-V-E-K-E-B-K.

FT REPEAT 1109 1116 1.

FT REPEAT 1117 1124 2.

FT SEQUENCE 1702 AA; 186539 MW; 860F70D2667807A6 CRC64;

Query Match 10.3%; Score 92.5; DB 1; Length 1702;

Best Local Similarity 23.9%; Pred. No. 44;

Matches 44; Conservative 26; Mismatches 101; Indels 13; Gaps 5;

QY 2 IVVKDFART---TWKEFLNNDGVEVSELKPHRVTVTIQNGKEMSTIVSE---EDPI 54

DB 1206 VVSKNQTEHTTQPTFERKTAIVETKQE--PQVSAQASPRQESSTVQVQVAVLESEN 1263

QY 55 LPVYKGELEKGYQPDGWEISFEGKKDA-GVYINLSKDTFIKPVFKIEKKEENKPTF 113

DB 1264 VPTVNAAEVQQLQGTQATVSTQKPAENISNTSATAIETAKSKPQETASTRE 1323

QY 114 DVSKKKDNQVNSHSHRDLQREHS---QKSDTKQVATVLDKNNTSSKSTTN 170

DB 1324 DASGKANTVADNSVANSSESEPKSRRRRSISQPETSAEETTAJSTDETTIADNSKSS 1383

QY 171 NPKK 174

DB 1384 KENR 1387

RESULT 13

YD72 SCHPO

ID YD72 SCHPO STANDARD; PRT; 1888 AA.

AC 014207;

DT 15-JUL-1998 (Rel. 36, Created)

DT 15-JUL-1998 (Rel. 36, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Hypothetical protein C6B12.02c in chromosome I.

GN SPAC6B12.02C.

OS Schizosaccharomyces pombe (fission yeast).

OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;

OC Schizosaccharomycetales; Schizosaccharomycetaceae;

OC Schizosaccharomyces.

OK NCBI\_Taxid:4896;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=972;

RX MEDLINE=21848401; PubMed=11859360;

RX Wood V., Gilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,

RA Sgouras J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,

RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,

RA Collins M., Connor R., Cronin A., Davis P., Feldwell T., Fraser A.,

RA Gentile S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,

RA Holroyd S., Hornby T., Howarth S., Huckle E.J., Hunt S., Jagsels K.,

RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,

RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odeh C.,

RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitch E.,

RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,

RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,

RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,

RA Woodward J., Wolchert G., Aert R., Robben J., Grumpprez B.,

RA Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,

RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,

RA Borzym K., Langer I., Beck A., Leinrich H., Reinhardt R., Pohl T.M.,

RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purrelle B.,

RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,

RA Gallbert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,

RA Lucas M., Rochet M., Galliard C., Tallada V.A., Garzon A., Thode G.,

RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey P., Benito J.,

RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forbury S.L.,

RA Cerruti L., Lowe T., McCombie M.R., Paulsen I., Potashkin J.,

RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;

RA "The genome sequence of Schizosaccharomyces pombe."

RT Nature 415:871-880(2002).

CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL, Z89531; CAB1064.1; -

DR PIR, T39009; T39009.

DR GeneDB, SPombe; SPAC6B12.02c; -

KW Hypothetical protein; Transmembrane.

FT TRANSMEM 697 717 POTENTIAL.

FT TRANSMEM 755 775 POTENTIAL.

FT TRANSMEM 866 886 POTENTIAL.

FT TRANSMEM 915 935 POTENTIAL.

FT TRANSMEM 977 997 POTENTIAL.

FT TRANSMEM 1033 1053 POTENTIAL.

FT TRANSMEM 1336 1356 POTENTIAL.

FT TRANSMEM 1645 1665 POTENTIAL.

FT DOMAIN 1662 1665 POLY-LEU.

FT SEQUENCE 1888 AA; 217432 MW; 8AD3BBCB3297C29 CRC64;

Query Match 10.3%; Score 92.5; DB 1; Length 1888;

Best Local Similarity 23.1%; Pred. No. 49;

Matches 39; Conservative 25; Mismatches 76; Indels 29; Gaps 5;

QY 11 TWKEFLNNDGVEVSELKPHRVTVTIQNGKEMSTIVSE-BDFILPVYKGELEKGYQPD 69

DB 394 TSENPFQLVANANASTIVYRTTKTKAKNRPKYVEKLPDLILESY----- 442

QY 70 GWEISFEGKKDAGVYINLSKDTFIKPVFKIEKKEENKPTFDVSKKKDNQVNSHSL 129

DB 443 -----GKKAPKPIKRVARSSHIP--KMIIRKQGMSSKKYFSPDXSDRQVIVDVS 492





```

OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C;
RX MEDLINE=95229044; PubMed=7713415;
RA Garsheoff L.J., Dollard C., Tan P., Winston F.;
RT "The Saccharomyces cerevisiae SPY7 gene encodes a very acidic protein
important for transcription in vivo.";
RL Genetics 139:523-536(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C;
RX MEDLINE=95076715; PubMed=7985423;
RA van der Aart Q.J.M., Barthe C., Doignon F., Aigle M., Crouzet M.,
Sleensma H.Y.;
RT "Sequence analysis of a 31 kb DNA fragment from the right arm of
Saccharomyces cerevisiae chromosome II.";
RL Yeast 10:959-964(1994).
RN [3]
RP SEQUENCE OF 1-835 FROM N.A.
RC STRAIN=S288C;
RA Andre B., Cziepluch C., Hein C., Jauniaux J.C., Urrestarazu A.,
Visseer S.;
RL Submitted (Aug-1994) to the EMBL/Genbank/DBJ databases.
RN [4]
RP SEQUENCE OF 463-523 FROM N.A.
RX MEDLINE=92285152; PubMed=1350857;
RA Haynes S.R., Dollard C., Winston F., Beck S., Trowdale J.,
David I.B.;
RT "The bromodomain: a conserved sequence found in human, Drosophila and
yeast proteins.";
RL Nucleic Acids Res. 20:2603-2603(1992).
CC -1- FUNCTION: TRANSCRIPTIONAL ACTIVATOR OF TY ELEMENTS AND POSSIBLY
OTHER GENES.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- SIMILARITY: Contains 1 bromodomain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L2537; AAC37424.1; -
DR EMBL; X76294; CAAS3940.1; -
DR EMBL; Z35950; CAAS5026.1; -
DR EMBL; M87651; AAA35087.1; -
DR PIR; S41552; S41552.
DR HSSP; Q92831; 1B91.
DR TRANSFAC; T04635; -
DR SGD; S0000285; SPY7.
DR GO; GO:000124; C:SAGA complex, IDA.
DR InterPro; IPR001487; Bromodomain.
DR Pfam; PF00439; bromodomain; 1.
DR PRINTS; PR00503; BROMODOMAIN.
DR SMART; SM00297; BROMO; 1.
DR PROSITE; PS00633; BROMODOMAIN 1; 1.
DR PROSITE; PS00634; BROMODOMAIN 2; 1.
KM Transcription regulation; Nuclear protein; Activator; Bromodomain.
FT DOMAIN 458 528 BROMODOMAIN.
FT SEQUENCE 1332 AA; 152616 MW; 083B63624669244F CRC64;
SQ
Query Match 10.3%; Score 92; DB 1; Length 1332;
Best Local Similarity 21.7%; Pred. No. 37; Mismatches 72; Gaps 12;
Matches 50; Conservative 32;
1 KIVVDFAFNTTVK-----EFI-----LNKDTGEVSELRPRV---- 33
Db 480 QIIKSMQNTLVTKLKSKFQYDSKQFVDIMLWNCNCLTNSDPHF--LKGHALIMQK 537
Qy 34 -----TVTIOGKEMSSIVSEEDFLLPVYKGLKGYQDPGWEISGFEKGKDG 83

```

```

Db 538 KSIQLIMIPITTRNADLKEI---EDM-----EKQDYELDEEEFVAGSRGK-- 585
Qy 84 YVIMLSKDTFIK--PVPKKIIEKKKEBKRPFD-----VSKKD-----NP 122
Db 586 --LWGAHMLAKENGKYSKSDSKTVDEATNDKLTSLVPEGKEKDKTAASTVTVHE 643
Qy 123 QVNHQSINSHRKEDLQF-BEHSQKSDSTKQVTVATVLD-KNNISKSTTN 170
Db 644 NVNANBEIKENKNEKQDWEESSTKEDSSKQADAKKQEDQLQKTAEN 693

RESULT 16
LIP STAB
ID LIP STAB STANDARD; PRT; 688 AA.
AC Q02510;
DT 01-JUN-1994 (Rel. 29, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Lipase precursor (EC 3.1.1.3) (Glycerol ester hydrolase).
GN GEHC OR SE0281.
OS Staphylococcus epidermidis.
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=1282;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 303-315.
RC STRAIN=9;
RX MEDLINE=93171870; PubMed=8436947;
RA Farrell A.M., Foster T.J., Holland K.T.;
RT "Molecular analysis and expression of the lipase of Staphylococcus
epidermidis.";
RL J. Gen. Microbiol. 139:267-277(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 12228;
RA Zhang Y., Ren S., Li H., Fu G., Lu L., Lu G., Jia J., Tu Y., Qin Z.,
Chen Z., Wen Y.;
RL Submitted (NOV-2002) to the EMBL/Genbank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: Triacylglycerol + a
fatty acid anion.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- MISCELLANEOUS: THE EXPRESSION OF STAPHYLOCOCCUS LIPASE IS
NEGATIVELY REGULATED BY BACTERIOPHAGE LYSOGENIZATION (LIPASE
CONVERSION).
CC -1- SIMILARITY: BELONGS TO THE ABHYDROLASE SUPERFAMILY. LIPASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M95577; AAA19729.1; -
DR EMBL; AB016744; AA003878.1; -
DR PIR; A47705; A47705.
DR InterPro; IPR005877; Gpos_YSIK.
DR InterPro; IPR000734; Lipase.
DR InterPro; IPR000379; Ser ester. site.
DR Pfam; PF04650; YSIK signal; 1.
DR TIGRPFAM; TIGR01168; YSIK signal; 1.
DR PROSITE; PS00120; LIPASE_SER; 1.
KM Hydrolase; Lipid degradation; Zymogen; signal; Complete proteome.
FT SIGNAL 1 35 POTENTIAL.
FT PROPEP 36 302 REMOVED IN THE MATURE FORM.
FT CHAIN 303 688 LIPASE.
FT ACT SITE 418 418 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT SITE 648 648 W -> L (IN REF. 1).
FT CONFLICT 96 96 E -> G (IN REF. 1).
FT CONFLICT 120 120 B -> G (IN REF. 1).
FT SEQUENCE 688 AA; 77343 MW; 6C95DB3A70AF66F6 CRC64;
SQ

```



```

FT      NP BIND      247      254      ATP (POTENTIAL) .
FT      SITE        352      355      DECH BOX.
SQ      SEQUENCE    778 AA;  88730 MM; 3B6C0857B5EABD84 CRC64;
Query Match      10.0%; Score 90; DB 1; Length 778;
Best Local Similarity 23.0%; Pred. No. 28;
Matches 44; Conservative 29; Mismatches 76; Indels 42; Gaps 6;

Qy      5 KDPARTTVAEPIAKDTGVESEIKPRVTTTTONGKEMST-----IVSEDPILP 56
      :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db      26 RNMNGVNVYKEKVENYDLTDISDSDSLDSEDNHGKGDNDTAPIMQLDDVHSDEDI-- 82
Qy      57 VTGKELEKQGFQDGMELISGFEKKDAGVINTLSQDTIKYVFKLIEKKEENKPTDVS 116
      83 -----QDSESDSDTEAVQ-AQVYKADKT--KSEQKSLDDELSEMDYKTVSLK 129
Qy      117 KKKDNPPQVNHSQLNES-----HKREDLQR-----EHSQKSDSTKVATVTL 158
      130 LKLTNEPVRQSYVSIIDTLTLHRSNEMVANANTKQNSNSDDEHSSKRRKTKKSIITDF 189
Qy      159 DKNNISSKSTT 169
Db      190 PKQKXKNEDTT 200

RESULT 18
TIG_STAM
ID      TIG_STAM      STANDARD;      PRT;      433 AA.
AC      Q09T16;
DT      28-FEB-2003 (Rel. 41, Created)
DT      28-FEB-2003 (Rel. 41, Last sequence update)
DT      28-FEB-2003 (Rel. 41, Last annotation update)
DS      Trigger factor (TF).
GN      TIG OR SAV1.675 OR SA1499 OR MM1619.
OS      Staphylococcus aureus (strain M50 / ATCC 700699) ,
OS      Staphylococcus aureus (strain N315), and
OS      Staphylococcus aureus (strain MW2).
OC      Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX      NCBI_TaxID=158878, 158879, 196620;
[1]
RN      RP
      SEQUENCE FROM N.A.
RC      MEDLINE=50 / ATCC 700699, and N315;
RX      MEDLINE=213195; PubMed=11418146;
RA      Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,
RA      Cui L., Oguchi A., Aoki K.-I., Nagai Y., Iian J.-O., Ito T.,
RA      Kanamori M., Matsuura H., Maruyama A., Murakami H., Hoshoyama A.,
RA      Mizutani-Ui Y., Takahashi N.K., Sawano T., Inoue R.-I., Kaito C.,
RA      Sekimizu K., Hiraekawa H., Kuhara S., Goto S., Yabuzaki J.,
RA      Kanehisa M., Yanaashita A., Oshima K., Furuya K., Yoshino C., Shiba T.,
RA      Hattori M., Ogasawara N., Hayashi H., Hiramatsu K.;
RT      "Whole genome sequencing of methicillin-resistant Staphylococcus
RT      aureus.";
RL      Lancet 357:1225-1240 (2001) .
[2]
RN      RP
      SEQUENCE FROM N.A.
RC      STRAIN=MM2;
RX      MEDLINE=22040717; PubMed=12044378;
RA      Baba T., Takeuchi F., Kuroda M., Yuzawa H., Aoki K.-I., Oguchi A.,
RA      Nagai Y., Iwano N., Asano K., Naimi T., Kuroda H., Cui L.,
RA      Yamanoto K., Hiramatsu K.;
RT      "Genome and virulence determinants of high virulence community-
RT      acquired MRSA.";
RL      Lancet 359:1819-1827(2002) .
[3]
RN      RP
      FUNCTION: Involved in protein export. Acts as a chaperone by
      maintaining the newly synthesized protein in an open conformation
      (by similarity).
[4]
RN      RP
      -1- SIMILARITY: BELONGS TO THE FKBP-TYPE PIASE FAMILY. TIG SUBFAMILY.
[5]
RN      RP
      This SWISS-PROT entry is copyright. It is produced through a collaboration
      between the Swiss Institute of Bioinformatics and the EMBL outstation -
      the European Bioinformatics Institute. There are no restrictions on its
      use by non-profit institutions as long as its content is in no way
      modified and this statement is not removed. Usage by and for commercial

```

```

RX MEDLINE=96310631; PubMed=8740425;
RA Sen-Gupta M., Lyck R., Fleg U., Niedenthal R.K., Hegemann J.H.;
RT "The sequence of a 24,152 bp segment from the left arm of chromosome
RT XIV from Saccharomyces cerevisiae between the BNI1 and the POL2
RT genes.";
RL Yeast 12:505-514(1996).
CC CC
CC -I- FUNCTION: ACTS ON PHOSPHATIDYLINOSITOL (PI) IN THE FIRST
CC COMMITTED STEP IN THE PRODUCTION OF THE SECOND MESSENGER
CC INSIGLIT-1,4,5,-TRISPHOSPHATE. PIK1 IS PART OF A NUCLEAR
CC PHOSPHOINOSITIDE CYCLE AND COULD CONTROL CYTOKINESIS THROUGH
CC THE ACTIN CYTOSKELETON.
CC CC
CC -I- CATALYTIC ACTIVITY: ATP + 1-phosphatidyl]-D-myo-inositol = ADP +
CC 1-phosphatidyl]-D-myo-inositol 4-phosphate.
CC CC
CC -I- SUBCELLULAR LOCATION: Nuclear.
CC CC
CC -I- SIMILARITY: BELONGS TO THE PI3/P14-KINASES FAMILY.
CC -----
CC CC
CC EMBL; X76058; CAAS3658.1; -.
CC DR EMBL; L20220; AAA34873.1; -.
CC DR EMBL; X92494; CAA63231.1; -.
CC DR EMBL; Z71543; CAA96174.1; -.
CC DR PIR; A49335; A49335.
CC SCD; S0005211; PIK1.
CC GO; GO:0005634; Cytoplasm; IDA.
CC GO; GO:0016288; P:cytoskeleton; IMP.
CC GO; GO:0006892; P:post Golgi transport; IGI.
CC InterPro; IPR000403; PI3_P14_kinase.
CC DR Pfam; PF00454; PI3_P14_kinase; 1.
CC SMART; SM00146; PI3kc; 1.
CC DR PROSITE; PS00915; PI3_4_KINASE_1; 1.
CC DR PROSITE; PS00916; PI3_4_KINASE_2; 1.
CC DR PROSITE; PS50290; PI3_4_KINASE_3; 1.
CC KW: Transferase; Kinase; Nuclear protein.
CC FT DOMAIN 793..1041 PI3K/PI4K.
CC SQ SEQUENCE 1066 AA; 119922 MW; 7666979CA14B1CB5 CRC64;

```

[1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-serotype D:  
 RA Wright A., Fisman Y., Tai F., Plant A.G.;  
 RL Submitted (May-1991) to the EMBL/GenBank/DBJ databases.  
 [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-Rd / KW20 / ATCC 51907;  
 RX MEDLINE=95350630; PubMed=7542800;  
 RA Pleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,  
 RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,  
 RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,  
 RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,  
 RA Weidman J.F., Phillips C.A., Spriggs T., Hedlom E., Cotton M.D.,  
 RA Usterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,  
 RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,  
 RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,  
 RA Venter J.C.;  
 RT "Whole-genome random sequencing and assembly of Haemophilus influenzae  
 Rd.";  
 RL Science 269:496-512(1995).  
 CC -I- FUNCTION: VIRULENCE FACTOR; CLEAVES HOST IMMUNOGLOBULIN A  
 CC PRODUCING INTRACT FC AND FAB FRAGMENTS.  
 CC -I- CATALYTIC ACTIVITY: Cleavage of immunoglobulin A molecules at  
 CC certain Pro-I-Xaa bonds in the hinge region. No small molecule  
 CC substrates are known.  
 CC -I- SUBCELLULAR LOCATION: Secreted.  
 CC -I- DOMAIN: THE SIGNAL PEPTIDE GUIDES THE PRECURSOR TO THE PERIPLASMIC  
 CC SPACE, AND THE CARBOXY-TERMINAL HELPER DOMAIN ASSOCIATES WITH THE  
 CC OUTER MEMBRANE TO FORM A PORE FOR EXCRETION OF THE PROTEASE  
 CC DOMAIN. THE HELPER DOMAIN IS THEN RELEASED BY AUTOROTOLYSIS (BY  
 CC SIMILARITY).  
 CC -I- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S6.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; X59800; -; NOT ANNOTATED\_CDS.  
 DR EMBL; U32779; AAC22651.1; -;  
 DR PIR; H64106; H64106.  
 DR MEROPS; S06.001; -;  
 DR TIGR; H10990; -;  
 DR InterPro; IPR006315; Autotransport.  
 DR InterPro; IPR005546; Autotransporter.  
 DR InterPro; IPR000710; IGA\_S6.  
 DR InterPro; IPR004899; Pertactin.  
 DR Pfam; PF03197; Autotransporter; 1.  
 DR Pfam; PF02395; IGA1; 1.  
 DR Pfam; PF03212; Pertactin; 1.  
 DR PRINTS; PR00921; IGASERPTASE.  
 DR TIGRFams; TIGR01414; autotrans\_bar1; 1.  
 DR HydroLase; Serine protease; Transmembrane; Zymogen; Signal;  
 KM HydroLase; Serine protease; Transmembrane; Zymogen; Signal;  
 KW Complete proteome.  
 FT SIGNAL 1 25 POTENTIAL.  
 FT CHAIN 26 1014 IMMUNOGLOBULIN A1 PROTEASE.  
 FT PROPEP 1015 1694 HELPER PEPTIDE (POTENTIAL).  
 FT ACT SITE 288 288 PROBABLE.  
 FT CONFLICT 253 254 BN -> GV (IN REF. 1).  
 FT CONFLICT 272 272 G -> A (IN REF. 1).  
 FT CONFLICT 464 464 G -> E (IN REF. 1).  
 FT CONFLICT 866 866 S -> T (IN REF. 1).  
 FT CONFLICT 1036 1036 A -> D (IN REF. 1).  
 FT CONFLICT 1074 1074 A -> G (IN REF. 1).  
 FT CONFLICT 1421 1421 A -> G (IN REF. 1).  
 FT CONFLICT 1545 1545 H -> T (IN REF. 1).  
 SQ SEQUENCE 1694 AA; 185539 MW; C52427013F93178C CRC64;

Query Match 10.0%; Score 89.5; DB 1; Length 1694;

Best Local Similarity 23.4%; Pred. No. 71;  
 Matches 43; Conservative 27; Mismatches 101; Indels 13; Gaps 5;  
 QY 2 IYKQPARNT---TYKEPIANKDGEVSEBLKHRTVTVIQNKEKSSSTVSE---EDFI 54  
 DB 1198 VSKNQTENTVTDQPTREKRTAKVETKTB--PPQVASQSPQOSESTVQPAVLESSEN 1255  
 QY 55 LPVYKHELEKQVQPMELISGPEKDA-GYVINLSKQFIKVPFKIEKKEBNKPTF 113  
 DB 1256 VPTVNNAEVQAQLQTSATVSTKQAPENSINTGSAATVETAKSDKPKQETAASTE 1315  
 QY 114 DYSKKKDNQVHNSQUNSHREDLDREHS---QKSDTKDTATVLDKNNISKSTTN 170  
 DB 1316 DSDQHNANTVADNSVANNSSSDPKSRRRRSISQPETSAEBTTASTDETTIADNSKRS 1375  
 QY 171 NPNK 174  
 DB 1376 KPNR 1379

## RESULT 21

YCG1\_YEAST STANDARD; PRT; 853 AA.  
 ID YCG1\_YEAST  
 AC P25588; P25589; P27513; P87003;  
 DT 01-MAY-1992 (Rel. 22, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE Hypothetical 97.9 kDa protein in CHAI-KR1 intergenic region.  
 GN YC1061C OR YC1061C/YC106C.  
 OS Saccharomyces cerevisiae (Baker's Yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.  
 OX NCBI\_TaxId=4932;  
 RN [1]

RP SEQUENCE FROM N.A.  
 RA Rasmussen S.W., von Wettstein D.;  
 RL Submitted (MAR-1992) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP REVISIONS.  
 RA Gromadka R.;  
 RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.  
 RN [3]

RP SEQUENCE OF 384-596 FROM N.A.  
 RC STRAIN=ATCC 28383 / FL100;  
 RX MEDLINE=90384830; PubMed=2169608;  
 RA Kern L.;  
 RT "The UKR1 gene of Saccharomyces cerevisiae encoding uridine kinase.";  
 RL Nucleic Acids Res. 18:5279-5279(1990).  
 CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL; X59720; CAA42405.1; -;  
 DR EMBL; X53998; -; NOT ANNOTATED\_CDS.  
 DR PIR; S74279; S74279.  
 DR SGD; S0000566; YC1061C.  
 DR GO; GO:0006347; P:chromatin silencing at HML and HMR (sensu S. .); IGI.  
 DR GO; GO:0006348; P:chromatin silencing at telomeres; IGI.  
 DR GO; GO:0000076; P:DNA replication checkpoint; IGI.  
 KW Hypothetical protein.  
 FT CONFLICT 505 505 L -> V (IN REF. 3).  
 FT CONFLICT 567 567 MISSING (IN REF. 3).  
 SQ SEQUENCE 853 AA; 97946 MW; 16809FCC08F248D1 CRC64;

Query Match 9.8%; Score 88; DB 1; Length 853;  
 Best Local Similarity 20.6%; Pred. No. 43;  
 Matches 40; Conservative 30; Mismatches 80; Indels 44; Gaps 7;

QY 4 VKDFANTTVKPEFLNKDGEVSELEKPHRVTVTIGNKEMSSSTIVSEEDFILPVY----- 58  
DB 266 LEDMAKKEIYENLENOEHI-----LRNKRI-----ROKREKREKLEENDFOJNAMDSSGD 305  
QY 59 KGELEKGYOPDGMWISGFEQ-----KKDAGYVINKSKDPFIKPFVKPIEK 104  
DB 306 SGSESGSGALSGNEIADYESSGSEENDNRESDEKEDDIIILKQKSHHVKKIINESD 365  
QY 105 KEENKPTFDVSKKKONPQVNSQINESHKEDLOREBHSQKSDSTKVTATVLDKNNI- 163  
DB 366 TEVEAKP-----KEKADESLPKRIAINLGHYGNI-----GDDTDKPFQETNVLDTQNI 414  
QY 164 ---SSKSTTNPNK 174  
DB 415 EYMAERNTEIENEVK 428  
RESULT 22  
FNR\_PEA STANDARD: PRT: 360 AA.  
ID\_FNR\_PEA  
AC P10933:  
DT 01-JUL-1989 (Rel. 11, Created)  
DT 01-JUL-1989 (Rel. 11, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Ferredoxin-NADP reductase, leaf isozyme, chloroplast precursor  
DE (BC 1.18.1.2) (FNR).  
GN PETH.  
OS Pisum sativum (Garden pea).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosids I; Fabales; Fabaceae; Papilionoideae; Viciae; Pisum.  
OX NCBI\_TaxId=3888;  
RN [1]  
RN SEQUENCE FROM N.A.  
RC STRAIN=cv. Little Marvel; TISSUE=leaf;  
RA Newman B.J., Gray J.C.;  
RT "Characterisation of a full-length cDNA clone for pea ferredoxin-NADP+  
RT reductase.";  
RL Plant Mol. Biol. 10:511-520(1988).  
RN [2]  
RN SEQUENCE OF 270-360 FROM N.A., AND MUTAGENESIS.  
RP MEDLINE=93374905; PubMed=8366077;  
RA Orellano E.G., Calcaterra N.B., Carrillo N., Ceccarelli E.A.;  
RA Orellano E.G., Calcaterra N.B., Ceccarelli E.A., Carrillo N.,  
RA Karplus P.A.;  
RT "A productive NADP+ binding mode of ferredoxin-NADP + reductase  
RT revealed by protein engineering and crystallographic studies.";  
RL Nat. Struct. Biol. 6:847-853(1999).  
CC -1- FUNCTION: MAY PLAY A KEY ROLE IN REGULATING THE RELATIVE AMOUNTS  
CC OF CYCLIC AND NON-CYCLIC ELECTRON FLOW TO MEET THE DEMANDS OF THE  
CC PLANT FOR ATP AND REDUCING POWER.  
CC -1- CATALYTIC ACTIVITY: Reduced ferredoxin + NADP(+) = oxidized  
CC ferredoxin + NADPH.  
CC -1- COFACTOR: FAD.  
CC -1- PATHWAY: FINAL STEP IN LINEAR PHOTOSYNTHETIC ELECTRON TRANSPORT  
CC CHAIN; IT HAS ALSO BEEN IMPLICATED IN CYCLIC ELECTRON FLOW AROUND  
CC PHOTOSYSTEM I. ITS ROLE BEING TO RETURN ELECTRONS FROM FERREDOXIN  
CC TO THE CYTOCHROME B-F COMPLEX.  
CC -1- SUBUNIT: Monomer.  
CC -1- SUBCELLULAR LOCATION: CHLOROPLAST; STROMAL SIDE OF THE THYLAKOID  
CC MEMBRANE IN THE VICINITY OF THE PHOTOSYSTEM I IN THE NON-STACKED  
CC AND FRINGE PORTION OF THE MEMBRANE.  
CC -1- MISCELLANEOUS: FNR IS PROBABLY ATTACHED TO THE MEMBRANE BY A  
CC SPECIFIC BINDING PROTEIN.  
CC -1- SIMILARITY: WITH OTHER SPECIES FNR.  
-----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL: X12446; CAA30978.1; -  
DR EMBL: L15565; AAB59349.1; -  
DR EMBL: L15567; AAB59303.1; -  
DR EMBL: L15569; AAB59304.1; -  
DR PIR: S04030; S04030.  
DR PDB: 1OPZ; 01-SEP-99.  
DR PDB: 1OPV; 01-SEP-99.  
DR PDB: 1OG0; 01-SEP-99.  
DR PDB: 1OGA; 01-SEP-99.  
DR InterPro: IPR001709; FNR\_cyc\_redtase.  
DR InterPro: IPR001433; Oxred\_FAD/NAD(P).  
DR Pfam: PF00175; NAD binding\_1; 1.  
DR PRINTS: PR00371; PFNCR.  
DR Oxidoreductase; Flavoprotein; NADP; FAD; Chloroplast; Transit peptide;  
KW Electron transport; Photosynthesis; Thylakoid; Membrane;  
KW Multigene family; 3D-structure.  
FT TRANSIT 1 52  
FT CHAIN 53 360  
FT DISULFID 178 183  
FT NP\_BIND 212 230  
FT MUTAGEN 360 360  
FT STRAND 68 68  
FT TURN 71 72  
FT STRAND 78 78  
FT TURN 80 81  
FT STRAND 84 93  
FT TURN 97 98  
FT STRAND 103 109  
FT TURN 111 112  
FT TURN 118 119  
FT STRAND 121 125  
FT STRAND 129 129  
FT TURN 131 132  
FT STRAND 135 135  
FT STRAND 139 143  
FT TURN 147 148  
FT TURN 150 151  
FT STRAND 156 162  
FT STRAND 165 167  
FT STRAND 169 170  
FT STRAND 173 175  
FT HELIX 177 184  
FT TURN 187 188  
FT STRAND 190 197  
FT TURN 200 201  
FT STRAND 204 204  
FT TURN 207 208  
FT STRAND 210 216  
FT HELIX 217 220  
FT HELIX 221 231  
FT TURN 232 232  
FT STRAND 243 250  
FT HELIX 253 255  
FT TURN 257 258  
FT HELIX 259 268  
FT TURN 270 272  
FT STRAND 273 279  
FT TURN 280 282  
FT STRAND 284 284  
FT TURN 286 287  
FT STRAND 290 290  
FT HELIX 293 297  
FT HELIX 298 300  
FT HELIX 301 309  
FT TURN 311 312

FT STRAND 313 319  
 FT TURN 321 322  
 FT HELIX 323 337  
 FT TURN 323 337  
 FT HELIX 338 339  
 FT TURN 342 351  
 FT TURN 352 353  
 FT STRAND 355 360  
 SQ SEQUENCE 360 AA; 40194 MW; 7PIC10DEBBA7B24 CRC64;

Query Match 9.8%; Score 87.5; DB 1; Length 360;  
 Best Local Similarity 24.6%; Pred. No. 19;  
 Matches 41; Conservative 21; Mismatches 54; Indels 51; Gaps 8;

QY 4 VKDFARNTT---VKEFILNKDTGAV-----SELRP-HRYVTYTGNGKEM----- 43  
 DB 148 IGFDFDSKTVSLCVARLVYTNDAEYVGVGVCNPLCDLKGSEGVKLTGPGKEMLPKDP 207  
 QY 44 -----SSTTVEEDFILPYVKGELKGYQPDG--WEISGFECKDAGVYNLSKDT 92  
 DB 208 NATVIMLGTCIAPFRSFLMKGF-FEKHEDYQFNGIAMLPLG-----VPTSSSL 256  
 QY 93 FIKPFPKLIBEKKERKPTPDVSKKDNPOVNHSQLNESHAKEDLQ 139  
 DB 257 LYKEFEFEKKEKAPENFILDFAVSRE-----QVNDKGEKMYIQ 294

RESULT 23  
 SKM1\_YEAST STANDARD; PRT; 655 AA.

ID SKM1\_YEAST STANDARD; PRT; 655 AA.  
 AC Q12469; Q06940;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Serine/threonine-protein kinase SKM1 (EC 2.7.1.1-) (Protein kinase 75490 D).  
 GN SKM1 OR YOL113W OR HRA655.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
 OX NCBI\_TaxID=4932;

RA Martin H., Mendoza A., Rodriguez-Pachon J.M., Molina M., Nombela C.;  
 "Characterization of SKM1, a Saccharomyces cerevisiae gene encoding a novel Ste20/Pdk-like protein kinase";  
 RL Mol. Microbiol. 23:431-444(1997).  
 RN [1]

RP SEQUENCE FROM N.A.  
 RC STRAIN=5288C / GRF88;  
 RA MEDLINE=97197189; PubMed=9044278;  
 RT Vanderbol M., Durand P., Portetelle D., Hilger F.;  
 "Sequence analysis of a 44 kb DNA fragment of yeast chromosome XV including the Ty1-H3 retrotransposon, the sufl1(+) frameshift  
 RT suppressor gene for trna-Gly, the yeast transfer RNA-thr-1a and a  
 RT delta element";  
 RL Yeast 11:1069-1075(1995).  
 CC -1- FUNCTION: MAY BE INVOLVED IN CELLULAR SIGNALING OR CYTOSKELETAL  
 CC FUNCTIONS. MAY PLAY A ROLE IN MORPHOGENETIC CONTROL.  
 CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.  
 CC STR20 SUBFAMILY.  
 CC -1- SIMILARITY: Contains 1 CRIB domain.  
 CC -1- SIMILARITY: Contains 1 PH domain.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see http://www.isb-sib.ch/announce/  
 CC or send an email to license@sib-sib.ch).  
 CC EMBL; X69322; CAA49163.1; -

DR EMBL; Z48149; CAA68147.1; -  
 DR EMBL; Z74855; CAA99132.1; -  
 DR PIR; S51884; S51884.  
 DR SGD; S0005473; SKM1.  
 DR GO; GO:0000074; P:regulation of cell cycle, IGI.  
 DR InterPro; IPR000095; PAKBox/Rb-binding.  
 DR InterPro; IPR001849; PH.  
 DR InterPro; IPR000719; Prot\_kinase.  
 DR InterPro; IPR002290; Ser\_thr\_kinase.  
 DR InterPro; IPR001245; Tyr\_pkinase.  
 DR Pfam; PF00786; PBD; 1.  
 DR Pfam; PF00169; PH; 1.  
 DR Pfam; PF00069; pkinase; 1.  
 DR PRINTS; PR00109; TYRKINASE.  
 DR ProDom; PD000001; Prot\_kinase; 1.  
 DR SMART; SM00285; PBD; 1.  
 DR SMART; SM00233; PH; 1.  
 DR SMART; SM00220; S\_TKC; 1.  
 DR PROSITE; PS50108; CRIB; 1.  
 DR PROSITE; PS50003; PH DOMAIN; 1.  
 DR PROSITE; PS50011; PROTEIN KINASE DOM; 1.  
 DR PROSITE; PS00108; PROTEIN\_KINASE\_ST; 1.  
 KW Transferase; Serine/threonine-protein kinase; ATP-binding.  
 FT DOMAIN 3 118  
 FT DOMAIN 123 136  
 FT DOMAIN 360 639  
 FT NP\_BIND 366 374  
 FT BINDING 406 406  
 FT ACT\_SITE 507 507  
 FT ACT\_SITE 303 303  
 FT CONFLICT 306 306  
 FT CONFLICT 320 320  
 FT CONFLICT 322 322  
 FT CONFLICT 452 452  
 FT CONFLICT 468 468  
 SQ SEQUENCE 655 AA; 75331 MW; 1F6CB85FEF10D385 CRC64;

Query Match 9.8%; Score 87.5; DB 1; Length 655;  
 Best Local Similarity 20.5%; Pred. No. 35;  
 Matches 34; Conservative 36; Mismatches 57; Indels 39; Gaps 7;

QY 16 FLINKDTGEVSELRKRYVTYTGNGKESSTTVEEDFILPYVKGEL-----EKGYP 68  
 DB 209 YLNRKTSIK---RSVSRTRKRGK-----TDSILPYQSRLKPPPSDDYRF 255  
 QY 69 DGEWEISGR-EKKDAGVYNLSKDTFIKPYKLIBEKKER-----EKKPTFDVK 117  
 DB 256 TWIEDNKVREER-----VHVSKESTADSQTKQKQKQKQYQSHLRHDDNNSTFRPH 309  
 QY 118 KKDNPQV--NHSQLNESHKEDLOREHHSQKSDSTKDYATATLDRN 161  
 DB 310 LAPSPATKRNDSKTKMKEDLLEKNNDSNELLIMKKTVAIDVA 355

RESULT 24  
 GARP\_PLAUF STANDARD; PRT; 678 AA.  
 ID GARP\_PLAUF STANDARD; PRT; 678 AA.  
 AC P13816;  
 DT 01-JAN-1990 (Rel. 13, Created)  
 DT 01-JAN-1990 (Rel. 13, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE Glutamic acid-rich protein precursor.  
 DB Glutamic acid-rich protein precursor.  
 GN GARP.  
 OS Plasmodium falciparum (isolate FC27 / Papua New Guinea).  
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
 OX NCBI\_TaxID=5837;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE=89040048; PubMed=2903445;  
 RA Triglia T., Stahl H.-D., Crewther P.E., Silva A., Anders R.F.,  
 RA Kemp D.J.;  
 RT "Structure of a Plasmodium falciparum gene that encodes a glutamic  
 RT acid-rich protein (GARP).";

```

RL Mol. Biochem. Parasitol. 31:199-202(1988).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL, J03998; AAA29605.1; -.
CC PIR, A54514; A54514.
CC Repeat; Malaria; Antigen; Signal.
CC KW SIGNAL.
CC FT SIGNAL 1 25
CC FT CHAIN 26 678
CC FT DOMAIN 120 164 15 X 3 AA TANDEM REPEATS OF K-K-X.
CC FT DOMAIN 372 416 9 X APPROXIMATE TANDEM REPEATS.
CC FT DOMAIN 417 441 5 X APPROXIMATE TANDEM REPEATS.
CC FT DOMAIN 576 604 POLY-GLU.
CC FT DOMAIN 605 653 7 X APPROXIMATE TANDEM REPEATS.
CC FT DOMAIN 654 663 POLY-GLU.
CC FT SEQUENCE 678 AA; 80551 MW; 2A8F8560496E9E CRC64;
SQ
Query Match
Best Local Similarity 24.1%; Score 87.5; DB 1; Length 678;
Matches 46; Conservative 35; Mismatches 73; Indels 37; Gaps 10;
QY 5 KDFARNTVKEFLNKD---TGEV---SELKRVTVTQNGKMSSTIVSEE---DF 53
DB 25 KCFNSGLKNQNLNNSPDSITGRLLNTELEKNK-----DNSSEITLKEKEDK 78
QY 54 ILPVYKGELEKGYQFGMEISGEGKKDAGVYINLSKDTPIKVFKEIEKK-----EE 107
DB 79 VPTTSDNKLKNHNNN--BIS---SSTDEPTNIIIVADKNENSVDDKKKKKKKKKK 133
QY 108 EKKPTFDVSKKQDNPVNSQLNESHKEDLQREHSQ-----KSTQDVATVADKN 162
DB 134 EKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK 191
QY 163 ISSK--STNN 171
DB 192 LDEWVSEITN 202
Db
RESULT 25
RA50_PRRHO STANDARD; PRT; 879 AA.
ID RA50_PRRHO STANDARD; PRT; 879 AA.
AC 058687;
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE DNA double-strand break repair rad50 ATPase.
GN RAD50 OR PH029.
OS Pyrococcus horikoshii.
OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OC Pyrococcus.
OX NCBI_Taxid=53953;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=OT3;
RX MEDLINE=98344137; PubMed=9679194;
RA Kawarabayashi Y., Sawada M., Horikawa H., Haikawa Y., Hino Y.,
RA Yamamoto S., Sekine M., Baba S.-I., Kosugi H., Hosoyama A., Nagai Y.,
RA Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamiya M., Ohfuku Y.,
RA Funahashi T., Tanaka T., Kudoh Y., Yamazaki J., Kushiida N., Oguchi A.,
RA Aoki K.-I., Yoshizawa T., Nakamura Y., Robb F.T., Horikoshi K.,
RA Masuchi Y., Shizuya H., Kikuchi H.;
RT "Complete sequence and gene organization of the genome of a hyper-
RT thermophilic archaeobacterium, Pyrococcus horikoshii OT3.";
RL DNA Res. 5:55-76(1998).
CC -f- FUNCTION: Involved in DNA double-strand break repair (DSBR). The
CC rad50/mre11 complex possesses single-strand endonuclease activity
CC and ATP-dependent double-strand-specific exonuclease activity.

```

```

CC Rad50 provides an ATP-dependent control of mre11 by unwinding
CC and/or repositioning DNA ends into the mre11 active site (By
CC similarity).
CC -f- SUBUNIT: Forms a complex with mre11 (By similarity).
CC -f- SIMILARITY: BELONGS TO THE SMC FAMILY. RAD50 SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL, AP000004; BAA30025.1; -.
CC PIR, C71083; C71083.
CC HAMAP; MF_00449; -; 1.
CC InterPro; IPR003593; AAA_ATPase.
CC InterPro; IPR003439; ABC_transporter.
CC InterPro; IPR003405; SMC_C.
CC InterPro; IPR003395; SMC_N.
CC Pfam; PF04423; Rad50_zn_hook; 1.
CC Pfam; PF02483; SMC_C; 1.
CC Pfam; PF02463; SMC_N; 1.
CC SMART; SM00382; AAA; 1.
CC DNA repair; Hydrolyase; ATP-binding; Coiled coil; Complete proteome.
CC NP_BIND 30
CC FT DOMAIN 141 744 COILED COIL (POTENTIAL).
CC FT SEQUENCE 879 AA; 103673 MW; FD4E30FD1B8CB29 CRC64;
SQ
Query Match
Best Local Similarity 21.0%; Score 87.5; DB 1; Length 879;
Matches 30; Conservative 35; Mismatches 51; Indels 27; Gaps 3;
QY 4 VDFARNTVKEFLNKDNGEVSSELKRVTVTQNGKMSSTIVSEEDFILFVYKGELE 63
DB 551 LNDYKNESYKLEIDKAKELSEIDRLRLGFTIDLSGI-----RELE 598
QY 64 KGYQFGMEISGEGKKDAGVYINLSKDTPIKVFKEIEKKKKKKKKPTFDVSKKQDNPQ 123
DB 599 K-----FNKK-----YIEAKNAKEKLDLIESLADERELDKAFELAKIEDIE 643
QY 124 VNSQLNESHKEDLQREHSQK 146
DB 644 KYTSQNLNQLRKPKDQKKYEKRE 666
Db
RESULT 26
UN89_CABEL STANDARD; PRT; 6632 AA.
ID UN89_CABEL STANDARD; PRT; 6632 AA.
AC 001761; Q17362;
DT 15-SRP-2003 (Rel. 42, Last sequence update)
DT 15-SRP-2003 (Rel. 42, Last annotation update)
DE Muscle M-line assembly protein unc-89 (Uncoordinated protein 89).
GN UNC-89 OR CO9D1.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_Taxid=6239;
RN [1]
RP SEQUENCE FROM N.A., FUNCTION, AND TISSUE SPECIFICITY.
RC STRAIN=Bristol N2;
RX MEDLINE=96180278; PubMed=8603916;
RA Benian G.M., Tinley T.L., Tang X., Borodovsky M.;
RT "The Caenorhabditis elegans gene unc-89, required for muscle M-line
RT transduction domains.";
RL J. Cell Biol. 132:835-848(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
Du 2., Le T.T., Wilson R.;

```

RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP REVISIONS.  
RA Waterston R.  
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.  
CC -1- FUNCTION: Structural component of the muscle M-line. Myofibril  
CC lattice assembly begins with positional cues laid down in the  
CC basement membrane and muscle cell membrane. UNC-89 responds to  
CC these signals, localizes, and then participates in assembling an  
CC M-line.  
CC -1- TISSUE SPECIFICITY: Localizes to the middle of A-bands.  
CC -1- SIMILARITY: Contains 1 DBL-homology (DH) domain.  
CC -1- SIMILARITY: Contains 1 fibronectin type III domain.  
CC -1- SIMILARITY: Contains 49 immunoglobulin-like C2-type domains.  
CC -1- SIMILARITY: Contains 1 PH domain.  
CC -1- SIMILARITY: Contains 5 RCSD domains.  
CC -1- SIMILARITY: Contains 1 SH3 domain.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See [http://www.isb-sib.ch/announce/  
CC or\\_send\\_email\\_to\\_license@isb-sib.ch](http://www.isb-sib.ch/announce/or_send_email_to_license@isb-sib.ch)).  
CC -----  
DR EMBL: U33058; AAB00542.1; -  
DR EMBL: AF003131; AAB54132.2; -  
DR PDB: 1FHO; 20-DEC-00.  
DR WormPep: C09D1.1; CE30426.  
DR InterPro: IPR003961; FN\_III.  
DR InterPro: IPR007110; IG\_1like.  
DR InterPro: IPR003598; IG\_c2.  
DR InterPro: IPR003006; IG\_MHC.  
DR InterPro: IPR001849; PH.  
DR InterPro: IPR007850; RCSD.  
DR InterPro: IPR000219; RhogGEP.  
DR InterPro: IPR001452; SH3.  
DR Pfam: PF000047; fn3; 1.  
DR Pfam: PF000169; PH; 1.  
DR Pfam: PF05177; RCSD; 5.  
DR Pfam: PF00621; RhogGEP; 1.  
DR Pfam: PF00018; SH3; 1.  
DR SMART: SMO0408; IGc2; 23.  
DR SMART: SMO0325; RhogGEP; 1.  
DR SMART: SMO0326; SH3; 1.  
DR PROSITE: PS50010; DH 2; 1.  
DR PROSITE: PS50835; IG-LIKE; 49.  
DR PROSITE: PS50003; PH\_DOMAIN; 1.  
DR PROSITE: PS50002; SH3; 1.  
KW Muscle protein; Immunoglobulin domain; Repeat; SH3 domain;  
KW 3D-structure.  
FT DOMAIN 63 127 SH3.  
FT DOMAIN 152 330 DH.  
FT DOMAIN 342 498 PH.  
FT DOMAIN 547 633 IG-LIKE C2-TYPE 1.  
FT DOMAIN 648 736 IG-LIKE C2-TYPE 2.  
FT DOMAIN 748 838 IG-LIKE C2-TYPE 3.  
FT DOMAIN 946 1033 IG-LIKE C2-TYPE 4.  
FT DOMAIN 1044 1132 IG-LIKE C2-TYPE 5.  
FT DOMAIN 1140 1227 IG-LIKE C2-TYPE 6.  
FT DOMAIN 1272 1315 THR-RICH.  
FT DOMAIN 1375 1475 RCSD 1.  
FT DOMAIN 1479 1585 RCSD 2.  
FT DOMAIN 1597 1695 RCSD 3.  
FT DOMAIN 1700 1799 RCSD 4.  
FT DOMAIN 1800 1860 RCSD 5.  
FT DOMAIN 1982 2067 IG-LIKE C2-TYPE 7.  
FT DOMAIN 2071 2163 IG-LIKE C2-TYPE 8.  
FT DOMAIN 2171 2261 IG-LIKE C2-TYPE 9.  
FT DOMAIN 2269 2359 IG-LIKE C2-TYPE 10.  
FT DOMAIN 2367 2455 IG-LIKE C2-TYPE 11.

FT	DOMAIN	2463	2564	IG-LIKE C2-TYPE 12.
FT	DOMAIN	2563	2651	IG-LIKE C2-TYPE 13.
FT	DOMAIN	2657	2746	IG-LIKE C2-TYPE 14.
FT	DOMAIN	2754	2858	IG-LIKE C2-TYPE 15.
FT	DOMAIN	2887	2980	IG-LIKE C2-TYPE 16.
FT	DOMAIN	2994	3081	IG-LIKE C2-TYPE 17.
FT	DOMAIN	3087	3183	IG-LIKE C2-TYPE 18.
FT	DOMAIN	3189	3280	IG-LIKE C2-TYPE 19.
FT	DOMAIN	3286	3376	IG-LIKE C2-TYPE 20.
FT	DOMAIN	3384	3469	IG-LIKE C2-TYPE 21.
FT	DOMAIN	3482	3572	IG-LIKE C2-TYPE 22.
FT	DOMAIN	3580	3667	IG-LIKE C2-TYPE 23.
FT	DOMAIN	3686	3777	IG-LIKE C2-TYPE 24.
FT	DOMAIN	3817	3908	IG-LIKE C2-TYPE 25.
FT	DOMAIN	3920	4009	IG-LIKE C2-TYPE 26.
FT	DOMAIN	4018	4106	IG-LIKE C2-TYPE 27.
FT	DOMAIN	4109	4201	IG-LIKE C2-TYPE 28.
FT	DOMAIN	4212	4297	IG-LIKE C2-TYPE 29.
FT	DOMAIN	4302	4387	IG-LIKE C2-TYPE 30.
FT	DOMAIN	4400	4485	IG-LIKE C2-TYPE 31.
FT	DOMAIN	4489	4580	IG-LIKE C2-TYPE 32.
FT	DOMAIN	4588	4678	IG-LIKE C2-TYPE 33.
FT	DOMAIN	4681	4771	IG-LIKE C2-TYPE 34.
FT	DOMAIN	4873	4961	IG-LIKE C2-TYPE 35.
FT	DOMAIN	4965	5057	IG-LIKE C2-TYPE 36.
FT	DOMAIN	5067	5160	IG-LIKE C2-TYPE 37.
FT	DOMAIN	5171	5260	IG-LIKE C2-TYPE 38.
FT	DOMAIN	5277	5366	IG-LIKE C2-TYPE 39.
FT	DOMAIN	5383	5472	IG-LIKE C2-TYPE 40.
FT	DOMAIN	5487	5578	IG-LIKE C2-TYPE 41.
FT	DOMAIN	5595	5685	IG-LIKE C2-TYPE 42.
FT	DOMAIN	5701	5790	IG-LIKE C2-TYPE 43.
FT	DOMAIN	5815	5904	IG-LIKE C2-TYPE 44.
FT	DOMAIN	5925	6014	IG-LIKE C2-TYPE 45.
FT	DOMAIN	6038	6130	IG-LIKE C2-TYPE 46.
FT	DOMAIN	6150	6239	IG-LIKE C2-TYPE 47.
FT	DOMAIN	6275	6368	FIBRONECTIN TYPE-III.
FT	DOMAIN	6413	6502	IG-LIKE C2-TYPE 48.
FT	DOMAIN	6507	6596	IG-LIKE C2-TYPE 49.
FT	DISULFID	568	621	POTENTIAL.
FT	DISULFID	2908	2975	POTENTIAL.
FT	DISULFID	3015	3065	POTENTIAL.
FT	DISULFID	3707	3759	POTENTIAL.
FT	DISULFID	3826	3890	POTENTIAL.
FT	DISULFID	5092	5157	POTENTIAL.
FT	DISULFID	5298	5350	POTENTIAL.
FT	DISULFID	5508	5560	POTENTIAL.
FT	DISULFID	5616	5669	POTENTIAL.
FT	DISULFID	5722	5764	POTENTIAL.
FT	DISULFID	5836	5901	POTENTIAL.
FT	DISULFID	5946	5998	POTENTIAL.
FT	DISULFID	6036	6171	POTENTIAL.
FT	DISULFID	6421	6486	POTENTIAL.
FT	CONFLICT	2137	2137	A -> P (IN REF. 1).
FT	CONFLICT	2245	2247	AKA -> PKP (IN REF. 1).
FT	CONFLICT	2258	2258	A -> P (IN REF. 1).
FT	CONFLICT	2284	2284	E -> G (IN REF. 1).
FT	CONFLICT	2297	2297	M -> I (IN REF. 1).
FT	CONFLICT	3531	3531	A -> G (IN REF. 1).
FT	CONFLICT	3884	3888	DAGEY -> RRRRI (IN REF. 1).
FT	CONFLICT	3929	3929	A -> V (IN REF. 1).
FT	CONFLICT	5134	5134	A -> P (IN REF. 1).
FT	CONFLICT	5145	5145	T -> S (IN REF. 1).
FT	CONFLICT	5185	5185	G -> A (IN REF. 1).
FT	CONFLICT	5199	5199	K -> N (IN REF. 1).
FT	CONFLICT	5202	5202	L -> P (IN REF. 1).
FT	CONFLICT	5213	5213	F -> L (IN REF. 1).
FT	CONFLICT	6178	6178	A -> G (IN REF. 1).
FT	CONFLICT	6268	6268	K -> E (IN REF. 1).
SO	SEQUENCE	6632 AA;	731665 MW;	262D3BDD6296089 CRC64;

Query Match Best Local Similarity

9.7%; 21.8%;

Score 87; DB 1; Length 6632; Pred. No. 4.7e+02;

Matches 37; Conservative 33; Mismatches 80; Indels 20; Gaps 7;

QY 11 TTVEFLINKDGEVSEBKPHRVVTIIONKEMSTIYVEEDP--ILPYKGELEKGYQF 68  
 DB 1127 TKAULTVQKPEPAPEDLPK--VSLTVEKGE--AVTSAAHPGIPLPYEMSVNGRKR 1181  
 QY 69 DQME---ISGFEKGKAGYVIMLSKDTPIKPY----FKKIEKK--EENKPTFDVSKKK 119  
 DB 1182 DQGEARVTRDSTYDGAISLTITDRTATYSEVNHLLTISVAENTLGAETGQQLTPEPK 1241  
 QY 120 DNPQVNHQSOLNESHKREDLQREHSGKSDSTQVATVLDKNNISSKSTT 169  
 DB 1242 BSVVEKQDPLSS-----EVQKHAQVVKASPEATTITMETSLSSTKTT 1287

RESULT 27  
 RAT1\_YEAST  
 ID RAT1\_YEAST STANDARD; PRT; 1006 AA.  
 AC 002752;  
 DT 01-OCT-1993 (Rel. 27, Created)  
 DT 01-OCT-1993 (Rel. 27, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Ribonucleic acid trafficking protein 1 (5'-3' exoribonuclease)  
 DE (EC 3.1.11.-) (P116).  
 GN RAT1 OR HKI1 OR TAP1 OR YOR048C.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 CC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
 OX NCBI\_TaxId=4932;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=92331925; PubMed=162825;  
 RA Amberg D.C., Goldstein A.L., Cole C.N.;  
 RT "Isolation and characterization of Rat1: an essential gene of  
 RT trafficking of mRNA.";  
 RT Genes Dev. 6:1173-1189 (1992).  
 RL [2]  
 RN [2]  
 RP SEQUENCE FROM N.A., AND EXORIBONUCLEASE ACTIVITY.  
 RX MEDLINE=93109318; PubMed=8417335;  
 RA Kenna M., Stevens A., McCammon M., Douglas M.G.;  
 RT "An essential yeast gene with homology to the exonuclease-encoding  
 RT XRN1/KEN1 gene also encodes a protein with exoribonuclease  
 RT activity.";  
 RT Mol. Cell. Biol. 13:341-350 (1993).  
 RL [3]  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=93368292; PubMed=8497260;  
 RA Aldrich T.L., di Segni G., McConaughy B.L., Keen N.J., Whelen S.,  
 RA Hall B.D.;  
 RT "Structure of the yeast TAP1 protein: dependence of transcription  
 RT activation on the DNA context of the target gene.";  
 RT Mol. Cell. Biol. 13:3434-3444 (1993).  
 RL [4]  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RX Landt O., Hiesel R., Unseld M.;  
 RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RX Bohn C., Bolotin-Fukuhara M., Daigman-Fornier B., Dang D.V.,  
 RA Valens M.;  
 RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: May function in the processing and/or trafficking of  
 CC nuclear mRNA. May be involved in general transcription as well.  
 CC Possesses 5'-3' exoribonuclease activity degrading poly(A) to  
 CC mainly 5'-AMP.  
 CC -1- COFACTOR: Requires magnesium.  
 CC -1- SUBCELLULAR LOCATION: Nuclear.  
 CC -1- SIMILARITY: Belongs to the 5'-3' exonuclease family.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; S61567; AAB26818.1; -  
 DR EMBL; M95626; AAA34960.1; -  
 DR EMBL; L06011; AAA16950.1; -  
 DR EMBL; Z11746; -; NOT ANNOTATED\_CDS.  
 DR EMBL; Z74956; CAA99240.1; -  
 DR PIR; S20126; S20126.  
 DR SGD; S0005574; RAT1.  
 DR GO; GO:0005634; C:nucleus; IDA.  
 DR GO; GO:004534; P:5'-3' exoribonuclease activity; IDA.  
 DR GO; GO:0006365; P:35S primary transcript processing; IMP.  
 DR GO; GO:0006396; P:RNA processing; IMP.  
 DR InterPro; IPR004859; Puc\_53exo.  
 DR Pfam; PF03159; XRN\_N; 1.  
 KW Nuclear protein; Hydroxylase; Nuclease; Exonuclease; Repeat.  
 FT DOMAIN 525 528 POLY-GLU.  
 FT DOMAIN 793 797 POLY-ASN.  
 FT DOMAIN 955 999 CONTAINS 2 X SRVD, 2 X NNNY, AND 2 X  
 FT YSGN REPEATS.  
 FT MUTAGEN 683 683 Y->H: IN ALLELE TAP1-1; ACTIVATES  
 FT TRANSCRIPTION OF THE PROMOTER-DEFECTIVE  
 FT YEAST SUP4 TRNA(TYR) ALLELE SUP4A53161.  
 FT SQ SEQUENCE 1006 AA; 115933 MW; 5DDDB0245F3E12A CRC64;

Query Match 9.6%; Score 86.5; DB 1; Length 1006;  
 Best Local Similarity 19.0%; Pred. No. 66;  
 Matches 32; Conservative 40; Mismatches 47; Indels 49; Gaps 8;

QY 37 IONGKEMSTIYSEBDFILPVY--KGELEKGYQFDGMEISGFGKDGAGYVIMLSKDTPI 94  
 DB 415 NKGQDRHHTVATEQ--LQMTDQGNLAKG-----SMNLTTSD-----WVRLKKEHML 460  
 QY 95 -----KVPFKIEKKEEENKPTFDVSKKDNQVNHSG-----LMSHRK 135  
 DB 461 ANEGNEEALAKYKQGDKNNEMLKDISKEIDDAVSKAMTNPFLAEVMQKILNKGRLL 520  
 QY 136 EDLQREH-----SOKSDSTQVATVLDK-----NNISSKSTT 169  
 DB 521 EKDNDEEETAKDSKKVTEKASECDLDAEIKDEIVADVDRNSBT 568

RESULT 28  
 RBP2\_PLAVB  
 ID RBP2\_PLAVB STANDARD; PRT; 1251 AA.  
 AC 000759;  
 DT 01-APR-1993 (Rel. 25, Created)  
 DT 01-APR-1993 (Rel. 25, Last sequence update)  
 DT 01-OCT-1996 (Rel. 34, Last annotation update)  
 DE Reticulocyte binding protein 2 (Fragment).  
 GN RBP2.  
 OS Plasmodium vivax (strain Belem).  
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
 OX NCBI\_TaxId=31273;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=92315338; PubMed=1617731;  
 RA Galinski M.R., Medina C.C., Ingravallo P., Barnwell J.W.;  
 RT "A reticulocyte-binding protein complex of Plasmodium vivax  
 RT merozoites.";  
 RT Cell 69:1213-1226 (1992).  
 RL Cell 69:1213-1226 (1992).  
 CC -1- FUNCTION: INVOLVED IN RETICULOCYTE ADHESION. SPECIFICALLY BINDS TO  
 CC HUMAN RETICULOCYTE CELLS.  
 CC -1- SUBCELLULAR LOCATION: Membrane-bound (Probable).  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial



CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL; M88098; AAA29744.1; --

KM Malaria; Receptor; Membrane.

FT NON\_TER 1

FT MON\_TER 1251 1251

SQ SEQUENCE 1251 AA; 143741 MW; 548A51C7404AC572 CRC64;

Query Match 9.6%; Score 86.5; DB 1; Length 1251;

Best Local Similarity 22.8%; Pred. No. 84;

Matches 46; Conservative 40; Mismatches 77; Indels 39; Gaps 10;

QY 2 IVVDFARNTYKRF-----ILKND--TGEVSEIKKH--RYVTYIIONGKEMSTI--- 47

DB 234 IALDAQIDTEVSKIEQINREINMKKEIKSYLSEIKYKCKCTTEISNRKGDKEFL 293

QY 48 -----VSEDFILPYKGELEKGYQFDGWEISGFE-GKKDAGYVNL--SKDTFIKVPK 100

DB 294 EKPRNESSNNKNNININININSEQYLDIEDAEKQASTKVELFKHETTSNFK 353

QY 101 -----IEKKKEENKPFEDVSK--KKNPOV-----NHSQINSHRKEDLOREBS 144

DB 354 SEILGVETTSQKINKAKEDIMKEIERHNSIQTQYKGFQENLKNLEPHNYDNABDELIN 413

QY 145 QKSDSTKDVATVLD--KNIS 164

DB 414 DKSTNAKVIETINLESVKHNS 435

RESULT 29

SYST\_LYCES

ID SYST\_LYCES STANDARD; PRT; 200 AA.

AC P27058;

DT 01-AUG-1992 (Rel. 23, Created)

DT 01-AUG-1992 (Rel. 23, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Systemin precursor.

OS Lycopersicon esculentum (Tomato).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

OC Asteridae; lamids; Solanales; Solanaceae; Solanum.

OX NCBI\_TaxId=4081;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Leaf;

RX MEDLINE=92196587; PubMed=1549783;

RA McGurl B., Pearce G., Orozco-Cardenas M., Ryan C.A.;

RT "Structure, expression, and antisense inhibition of the systemin precursor gene.";

RL Science 255:1570-1573(1992).

RN [2]

RP SEQUENCE OF 179-196.

RA Pearce G., Striydom D., Johnson S., Ryan C.A.;

RT "A polypeptide from tomato leaves induces wound-inducible proteinase inhibitor proteins.";

RL Science 253:895-898(1991).

RN [3]

RP FUNCTION.

RX MEDLINE=20175266; PubMed=10708853;

RA Ryan C.A.;

RT "The systemin signaling pathway: differential activation of plant defensive genes.";

RL Biochem. Biophys. Acta 1477:112-121(2000).

-1- FUNCTION: ACTIVATES A LIPID-BASED SIGNAL TRANSDUCTION PATHWAY IN WHICH LINOLENIC ACID IS CONVERTED TO JASMONIC ACID, A POTENT ACTIVATOR OF DEFENSE GENE TRANSCRIPTION, INCLUDING PROTEINASE INHIBITOR.

-1- SUBCELLULAR LOCATION: Cytoplasmic.

-1- TISSUE SPECIFICITY: ALL ORGANS EXCEPT THE ROOTS. TRANSPORTED OUT OF WOUNDS TO DISTAL TISSUES.

-1- INDUCTION: By wounding; in leaves.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - CC the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL; M84800; AAA34182.1; --

DR EMBL; M84801; AAA34184.1; --

DR PIR; T07149; T07149.

KM Hormone; Repeat.

FT PROPEP 1 178

FT CHAIN 179 196

FT PROPEP 197 200

FT REPEAT 3 8

FT REPEAT 37 45

FT REPEAT 80 88

FT REPEAT 117 125

FT REPEAT 145 153

SQ SEQUENCE 200 AA; 22999 MW; FA251D94BAA9C5A9 CRC64;

Query Match 9.6%; Score 86; DB 1; Length 200;

Best Local Similarity 21.5%; Pred. No. 13;

Matches 37; Conservative 30; Mismatches 59; Indels 46; Gaps 8;

QY 1 KIIVK-----DFANTYKERILKNDGVESEIKPHRTVYIIONGKEMSTIVSEDFIL 55

DB 34 KIIEKTPSQDINKKDDTSSVLRDDPQELPKM-H-----EEGGYKKEIKVETTSQ 86

QY 56 PVYKGELEKGYQFDGWEISGFEKKDAGYVNL--SKDTFIKVPKIEBKE--EENKPT 112

DB 87 YIIR-----IEGDDA-----QETKVEYEEBEKKEIKVETTSQ 122

QY 113 FDVSKKDN-----PQVNHQSQINSHRKEDLOREBSQKSDSTKDVATVLDK 160

DB 123 QDINKKDDDGKPKVEH---EAGDDKETPSQDIIKKGEGALRTTKVCEK 171

RESULT 30

GLNA\_PYRKO

ID GLNA\_PYRKO STANDARD; PRT; 443 AA.

AC O08467;

DT 15-DEC-1998 (Rel. 37, Created)

DT 15-DEC-1998 (Rel. 37, Last sequence update)

DT 15-DEC-1998 (Rel. 37, Last annotation update)

DE Glutamine synthetase (GS).

OS Pyrococcus kodakarensis.

OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;

OC Thermococcus.

OX NCBI\_TaxId=69014;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=KOD1;

RX MEDLINE=97316461; PubMed=9172372;

RA Adul Rahman R.N.Z., Tongaareejit B., Fujiwara S., Imanaka T.;

RT "Characterization of recombinant glutamine synthetase from the hyperthermophilic archaeon Pyrococcus sp. strain KOD1.";

RL Appl. Environ. Microbiol. 63:2472-2476(1997).

CC -1- CATALYTIC ACTIVITY: ATP + L-glutamate + NH(3) = ADP + phosphate + L-glutamine.

CC -1- SUBCELLULAR LOCATION: Cytoplasmic (potential).

CC -1- SIMILARITY: BELONGS TO THE GLUTAMINE SYNTHETASE FAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - CC the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: D86222; BAA20530.1; -  
 DR HSPSP; P06201; ILGR.  
 DR InterPro; IPR001691; GLN\_synth.  
 DR InterPro; IPR004809; GLNA.  
 DR InterPro; IPR001637; GLNA\_adenyltn.  
 DR Pfam; PF03951; gln-synt\_N; 1.  
 DR Pfam; PF03951; gln-synt\_C; 1.  
 DR ProDom; PD001057; gln-synt\_N; 1.  
 DR TIGRFam; TIGR00653; GLNA; 1.  
 DR PROSITE; PS00180; GLNA\_1; 1.  
 DR PROSITE; PS00181; GLNA\_ATP; 1.  
 KM LIGASE.  
 FT BINDING.  
 FT 362 362 AMP (UNDER CONDITIONS OF ABUNDANT  
 FT SEQUENCE 443 AA; 50259 MW; 9426DCFE8F18168 CRC64;  
 FT  
 SQ  
 Query Match 9.6%; Score 86; DB 1; Length 443;  
 Best Local Similarity 17.8%; Pred. No. 30;  
 Matches 31; Conservative 25; Mismatches 46; Indels 72; Gaps 4;

QY 58 YKGELEKGVQPDGWMISFGFGKKDAGVYNLSKDPFI----- 94  
 DB 42 YEEAVEDGVSPFGSSIPFGEGIEDSLFKADPSTYALIPWEGIGRVGYIKGDEPYQA 101  
 QY 95 --KPVFKKJBKKKEEN-----KPTFVSKKD----- 120  
 DB 102 DPGILKRVLEKRGKAKAIGPEDEFIFKKNGTWELHIDSGGYFDLVGDKARER 161  
 QY 121 -----NPOVNSQLNESHRKEDLOREHSQKSD---STQDVATATVD 159  
 DB 162 REIALYMPYLGLKPEVLHHRGKAGQHEIDFRYDEALRTADNIVSPGHVAKAAE 215

RESULT 31  
 PBPA\_BACSU STANDARD; PRT; 914 AA.  
 AC P39793;  
 DT 01-FEB-1995 (Rel. 31, Created)  
 DT 01-FEB-1995 (Rel. 31, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Penicillin-binding protein 1A/1B (PBp1) [includes: Penicillin-  
 DE insensitive transglycosylase (EC 2.4.2.-) (peptidoglycan Tgase);  
 DE penicillin-sensitive transpeptidase (EC 3.4.-.-) (PB-transpeptidase)].  
 GN PONA.  
 OS Bacillus subtilis.  
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
 OX NCBI\_TaxID=1423;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 499-515.  
 RC STRAIN=168;  
 RX MEDLINE=95113769; PubMed=7814321;  
 RA Popham D.L., Setlow P.;  
 RT "Cloning, nucleotide sequence, and mutagenesis of the Bacillus  
 RT (PBp) 1 and a PBP-related factor.";  
 RL J. Bacteriol. 177:326-335 (1995).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=168 / Marburg;  
 RX MEDLINE=96349105; PubMed=8760912;  
 RA Sorokin A.V., Azevedo V., Zumbstein E., Galleron N., Ehrlich S.D.,  
 RA Serrro P.;  
 RT "Sequence analysis of the Bacillus subtilis chromosome region between  
 RT the *serA* and *kds* loci cloned in a yeast artificial chromosome.";  
 RL Microbiology 142:2005-2016 (1996).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=168;  
 RX MEDLINE=98044033; PubMed=9384377;  
 RA Kunert P., Ogasawara N., Mosser I., Albertini A.M., Allion G.,  
 RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,  
 RA Boriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,  
 RA Brouillet S., Brusch C.V., Caldwell B., Capiano V., Carter N.M.,

RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,  
 RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,  
 RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,  
 RA Fritz C., Fujita M., Fujita Y., Fuma S., Gallazzi A., Galleron N.,  
 RA Ghim S.Y., Glaser P., Goffeau A., Gollightly E.J., Grand G.,  
 RA Guisepert G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,  
 RA Hilbert H., Holstappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,  
 RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,  
 RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,  
 RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,  
 RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,  
 RA Medina N., Mellado R.P., Mizuno M., Mestl D., Nakai S., Noback M.,  
 RA Noone D., O'Reilly M., Ogawa K., Ogawara A., Oudega B., Park S.H.,  
 RA Parro V., Pohl T.M., Portetelle D., Potwilk S., Prescott A.M.,  
 RA Priescan E., Pujic P., Putrelle B., Rapoport G., Ray M., Reynolds S.,  
 RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadale Y.,  
 RA Sato T., Scanlan E., Schleich S., Schoeter R., Scoffone P.,  
 RA Sekiguchi J., Sekowska A., Serrro S.J., Serrro P., Shin B.S., Soldo B.,  
 RA Sorokin A., Taccioni B., Takagi T., Takahashi H., Takemura K.,  
 RA Takeuchi M., Tamakoshi A., Tanaka T., Terpsate P., Tognoni A.,  
 RA Tosato V., Uchiyama S., Vandenbol M., Vannier P., Vassaretti A.,  
 RA Viati A., Wambut R., Wedler E., Wedler H., Wetzenecker T.,  
 RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,  
 RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.,  
 RT "The complete genome sequence of the Gram-positive bacterium Bacillus  
 RT subtilis.";  
 RL Nature 390:249-256 (1997).  
 RN [4]  
 RP GROWTH REQUIREMENTS.  
 RC STRAIN=168;  
 RX MEDLINE=98389671; PubMed=9721295;  
 RA Sorokin A., Popham D.L., Setlow P.;  
 RT "Bacillus subtilis cells lacking penicillin-binding protein 1 require  
 RT increased levels of divalent cations for growth.";  
 RL J. Bacteriol. 180:4555-4563 (1998).  
 RN [5]  
 RP SUBCELLULAR LOCATION.  
 RC STRAIN=168;  
 RX MEDLINE=99255546; PubMed=10322023;  
 RA Pedersen L.B., Angert E.R., Setlow P.;  
 RT "Septal localization of penicillin-binding protein 1 in Bacillus  
 RT subtilis.";  
 RL J. Bacteriol. 181:3201-3211 (1999).  
 CC -1- FUNCTION: CELL WALL FORMATION. SYNTHESIS OF CROSS-LINKED  
 CC PEPTIDOGLYCAN FROM THE LIPID INTERMEDIATES. THE ENZYME HAS A  
 CC PENICILLIN-INSENSITIVE TRANSGLYCOSYLASE N-TERMINAL DOMAIN  
 CC (FORMATION OF LINEAR GLYCAN STRANDS) AND A PENICILLIN-SENSITIVE  
 CC TRANSPEPTIDASE C-TERMINAL DOMAIN (CROSS-LINKING OF THE PEPTIDE  
 CC SUBUNITS) (BY SIMILARITY).  
 CC -1- PATHWAY: Peptidoglycan synthesis; final stages.  
 CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. PROBABLY FOUND ALL  
 CC OVER THE WHOLE CELL AT LOW CONCENTRATIONS. ALSO LOCALIZES TO THE  
 CC DIVISION SITE IN VEGETATIVE CELLS.  
 CC -1- DEVELOPMENTAL STAGE: EXPRESSION IS CONSTANT DURING GROWTH,  
 CC DECREASES DURING SPOULATION AND IS INDUCED APPROXIMATELY 15 MIN  
 CC INTO SPORE GERMINATION.  
 CC -1- PTM: THE PRODUCT EXPRESSED FROM THE TRANSLATION OF THE PONA GENE  
 CC APPEARS AS TWO BANDS ON A GEL (1A AND 1B), BUT THE SPECIFIC AMINO  
 CC ACID SEQUENCE OF EACH PROTEIN IS UNKNOWN.  
 CC -1- PTM: THE N-TERMINUS IS BLOCKED.  
 CC -1- MISCELLANEOUS: CELLS LACKING THE PROTEIN REQUIRE INCREASED LEVELS  
 CC OF MG(2+) OR CA(2+) FOR GROWTH AND GERMINATION. APPROXIMATELY 50%  
 CC OF CELLS WITHOUT THE PROTEIN CONTAIN ABNORMAL FTZ RINGS,  
 CC SUGGESTING IT IS INVOLVED IN SEPTUM SYNTHESIS; INCREASED LEVELS OF  
 CC MG(2+) OR CA(2+) ONLY PARTIALLY ELIMINATE THE SEPTATION DEFECTS.  
 CC -1- SIMILARITY: contains 1 fibronectin type III domain.  
 CC -1- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE  
 CC TRANSGLYCOSYLASE FAMILY.  
 CC -1- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO THE  
 CC TRANSPEPTIDASE FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----

DR EMBL; U11883; AAA64947.1; -

DR EMBL; L47838; AAB38459.1; -

DR PIR; 299115; CAB14148.1; -

DR PIR; 140529; 140529.

DR Subtilisin; BGI0954; poa.

DR InterPro; IPR003961; FN\_III.

DR InterPro; IPR001264; Glyco. Trans. 51.

DR InterPro; IPR001460; Transpeptidase.

DR Pfam; PR00041; fn3; 1.

DR Pfam; PR00912; Transglycosyl; 1.

DR Pfam; PR00905; Transpeptidase; 1.

DR ProDom; PD001895; Glyco. Trans. 51; 1.

DR SMART; SM00060; FN3; 1.

KM Peptidoglycan synthetase; Cell wall; Transferase; Glycosyltransferase;

KM Hydroxase; Multifunctional enzyme; Transmembrane; Signal-anchor;

KM Antibiotic resistance; Complete proteome.

FT TRANSMEM 1 37 CYTOPLASMIC (POTENTIAL).

FT DOMAIN 38 58 SIGNAL-ANCHOR (TYPE II MEMBRANE PROTEIN)

FT DOMAIN 59 914 (POTENTIAL)

FT DOMAIN 77 246 EXTRACELLULAR (POTENTIAL).

FT DOMAIN 329 662 TRANSGLYCOSYLASE.

FT DOMAIN 706 784 TRANSPEPTIDASE.

FT ACT SITE 390 FIBROBLAST TYPE-III.

FT ACT SITE 390 ACYLATED BY PENCICLIN (BY SIMILARITY).

SO SEQUENCE 914 AA; 99562 MW; 697833DFE2423E6 CRC64;

Query Match 9.6%; Score 86; DB 1; Length 914;

Best Local Similarity 23.0%; Pred. No. 65;

Matches 45; Conservative 33; Mismatches 56; Indels 62; Gaps 9;

QY 13 VKEPLANDTGEVSLKPHRTVTITONGKEMSTVSEDFLLPYKGLKGYOPDQWE 72

DB 737 VKQSV---DGGSYSE-----IONS6-----AKKAVISGVQPGSV---YKREYVA 774

QY 73 ISGPGKADA--GYVINSKDTPIKPVFKIKKEKKEBK----- 110

DB 775 VSD--DGKSTASTSYEPKABDEDEKDDQQTDEKDDTDTQTDSDSQKDDGQTDQD 833

QY 111 ----FTFVSKKKKDPQVNHSQLN-----ESHKKDLQREHHSQKSDSTQVTA 155

DB 834 QTDSDTNDQKKQDNTNTPSDNNQDSDNDNDNNSNQDSDSDSGSKXDSTG--SD 891

QY 156 TVLDKNNISKSSTNN 171

DB 892 TNKNTDTSNKTQINS 907

RESULT 32

MDN1\_HUMAN STANDARD; PRT; 5596 AA.

AC Q9NU02; O15019;

DT 28-FEB-2003 (Rel. 41, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DB Midasin (MIDAS-containing protein).

GN MDN1 OR KIAA0301.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Mammalia; Eubacteria; Primates; Catarrhini; Homiidae; Homo.

OX NCBI\_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Testis;

RX PubMed=12102729;

RA Garbarino J.E., Gibbons I.R.;

RT "Expression and genomic analysis of midasin, a novel and highly conserved AAA protein distantly related to dynein.";

RL BMC Genomics 3:18-18(2002).

RN [2]

RP SEQUENCE OF 1255-2356 AND 3550-5596 FROM N.A.

RC TISSUE=Testis;

RA Tracey A.;

RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.

RN [3]

RP SEQUENCE OF 3550-5596 FROM N.A.

RC TISSUE=Brain;

RX MEDLINE=97349984; PubMed=9205841;

RA Nagase T., Ishikawa K.-I., Nakajima D., Ohira M., Seki N.,

RA Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;

RT Prediction of the coding sequences of unidentified human genes. VII.

RT The complete sequences of 100 new cDNA clones from brain which can

RT code for large proteins in vitro."

RL DNA Res. 4:141-150(1997).

CC -1- FUNCTION: May function as a nuclear chaperone and be involved in

CC the assembly/disassembly of macromolecular complexes in the

CC nucleus.

CC -1- SUBCELLULAR LOCATION: Nuclear (By similarity).

CC -1- SIMILARITY: Contains 1 VWA domain.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL; AF503925; AAM77722.1; -

DR EMBL; AL096678; CAB86660.1; -

DR EMBL; AL096678; CAB86660.1; -

DR EMBL; AB002299; BAA20761.1; -

DR Genew; HGNC:18302; MDN1.

DR GO; GO:0005634; C:nucleus; NAS.

DR GO; GO:0016887; F:ATPase activity; NAS.

DR GO; GO:0003754; F:chaperone activity; NAS.

DR GO; GO:0006461; P:protein complex assembly; NAS.

DR InterPro; IPR003593; AAA\_ATPase.

DR InterPro; IPR002035; VWF\_A.

DR Pfam; PF00004; AAA; 1.

DR SMART; SM00382; AAA; 7.

DR SMART; SM00327; VWA; 1.

DR PROSITE; PS0234; VWA; 1.

DR Chaperone; ATP-binding; Repeat; Nuclear protein.

KW Chaperone; ATP-binding; Repeat; Nuclear protein.

FT NP\_BIND 329 336 ATP (POTENTIAL).

FT NP\_BIND 677 684 ATP (POTENTIAL).

FT NP\_BIND 1084 1091 ATP (POTENTIAL).

FT NP\_BIND 1390 1397 ATP (POTENTIAL).

FT NP\_BIND 1753 1760 ATP (POTENTIAL).

FT NP\_BIND 2066 2073 ATP (POTENTIAL).

FT DOMAIN 3566 3573 POLY-GLU.

FT DOMAIN 4784 4791 POLY-GLU.

FT DOMAIN 5008 5013 POLY-GLU.

FT DOMAIN 5182 5187 POLY-GLU.

FT DOMAIN 5384 5583 POLY-GLU.

FT CONFLICT 2287 2312 RFLSMIDPVHGDISRAMNRNGLEIYI -> S (IN REF.

SO SEQUENCE 5596 AA; 632802 MW; 586C62616A1F9604 CRC64;

Query Match 9.6%; Score 86; DB 1; Length 5596;

Best Local Similarity 23.4%; Pred. No. 4,6e+02;

Matches 30; Conservative 27; Mismatches 63; Indels 8; Gaps 2;

QY 28 LKPHRYVTITONGKEMSTVSEDFLLPYKGLKGYOPDQWE 81

DB 4629 LATHRSTALISVLAQVFTELQAGFCFLPKERMEISAGCATEFHBYBGJIGEGGMKD 4688

QY 82 AGVAINLSKDTPIKPVFKIKKEKKEBKPTFVSKKKDPQVNHSQLNESHKKDLQRE 141

DB 4688 VS--DQIGNEQVEDTFOGQKDEKDEPDPSKSDIKGEDNAIEMSEDFDGMHGGLEBDE 4746

QY 142 EHSOKSDS 149  
DB 4747 EDEKSDS 4754

RESULT 33  
DRL1 YEAST STANDARD; PRT; 752 AA.  
AC P32892;  
DT 01-OCT-1993 (Rel. 27, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 15-SEP-2003 (Rel. 42, Last annotation update)  
DE Probable ATP-dependent RNA helicase DRL1.  
GN DRL1 OR YLL008M OR I1345.  
OS Saccharomyces cerevisiae (Baker's Yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.  
OX NCBI\_TaxID=4932;  
RN (1)  
RP SEQUENCE FROM N.A.  
RX MEDLINE=93087480; PubMed=1454790;  
RA Ripmaster T.L., Vaughn G.P., Woolford J.L., Jr.;  
RT "A putative ATP-dependent RNA helicase involved in Saccharomyces  
RT cerevisiae ribosome assembly.";  
RT Proc. Natl. Acad. Sci. U.S.A. 89:11131-11135 (1992).  
RL (2)  
RN SEQUENCE FROM N.A.  
RP STRAIN=S288C / FY23;  
RX MEDLINE=96405918; PubMed=8810043;  
RA Miosga T., Zimmermann F.K.;  
RT "Sequence analysis of the CEN12 region of Saccharomyces cerevisiae on  
RT a 4.7 kb fragment of chromosome XII including an open reading frame  
RT homologous to the human cystic fibrosis transmembrane conductance  
RT regulator protein CFTR.";  
RT Yeast 12:693-708 (1996).  
RL (3)  
RN SEQUENCE FROM N.A.  
RP STRAIN=S288C / AB972;  
RX MEDLINE=9711267; PubMed=9169871;  
RA Johnston M., Hillier L., Riles L., Albermann K., Andre B., Ansgore W.,  
RA Benes V., Brueckner M., Delius H., Dubois E., Duesterhoeft A.,  
RA Eutlian K.-D., Floeth M., Goffeau A., Hebling U., Heumann K.,  
RA Heuss-Nietzel D., Hilbert H., Hilger F., Kleine K., Koeltter P.,  
RA Louis E.J., Messenguy F., Mewes H.-W., Miosga T., Moestl D.,  
RA Mueller-Auer S., Neutwich U., Obermaier B., Pirvandi E., Pohl T.M.,  
RA Portetelle D., Putnelli B., Reckmann S., Rieger M., Rinke M., Rose M.,  
RA Scharte M., Scherens B., Scholler P., Schwager C., Schwarz S.,  
RA Underwood A.P., Urrestarazu L.A., Vandenbol M., Verhasselt P.,  
RA Vierendels F., Voet M., Volckaert G., Voss H., Wambutt R., Wedler B.,  
RA Wedler H., Zimmermann F.K., Zollner A., Hani J., Hohlweis J.D., XII.";  
RT "The nucleotide sequence of Saccharomyces cerevisiae chromosome XII.";  
RT Nature 387:87-90 (1997).  
RL (1)  
RN FUNCTION: PROBABLE HELICASE INVOLVED IN RIBOSOME ASSEMBLY.  
CC -1- SUBCELLULAR LOCATION: Nuclear.  
CC -1- SIMILARITY: BELONGS TO THE DEAD BOX HELICASE FAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: L00683; AAA34666.1; -;  
DR EMBL: X91488; CAA62783.1; -;  
DR EMBL: Z73113; CAA97452.1; -;  
DR PIR: S64750; S64750.  
DR HSSP: OS8083; 1HV8.  
DR SGD: S0003931; DRL1.  
DR InterPro: IPR001410; DEAD.  
DR InterPro: IPR000629; DEAD box.  
DR InterPro: IPR001650; Helicase\_C.

DR Pfam; PF00270; DEAD, 1.  
DR Pfam; PF00271; Helicase\_C; 1.  
DR SMART; SM00487; DEXDC; 1.  
DR SMART; SM00490; HELIC; 1.  
DR PROSITE; PS00039; DEAD\_ATP\_HELICASE; 1.  
KW ATP-binding; RNA-binding; Helicase; Nuclear protein.  
FT DOMAIN 170 190 POLY-GLU.  
FT NP\_BIND 275 282 ATP (POTENTIAL).  
FT SITE 385 388 DEAD BOX.  
FT CONFLICT 1 68 MYGTRKTSNLDVFPVTSSEDDVPIIDSSDDEKVKAKTT  
FT KRRGKNNKKVSEGDINDEVDHBDL -> MTKSRRLRL  
FT RRSGRVLRITKRLVIGTSMRPMRTMT (IN REF. 1).  
SQ SEQUENCE 752 AA; 84843 MW; 60747607A65B4A8 CRC64;  
Query Match 9.5%; Score 85.5; DB 1; Length 752;  
Best Local Similarity 21.8%; Pred. No. 57;  
Matches 38; Conservative 29; Mismatches 64; Indels 43; Gaps 8;  
QY 17 ILANKTGVESELKPRVYTTTNGKEMSTIYSEDFILPYVKGELKGYOPD----- 69  
DB 26 IIDSDDDEKVEAKR--TTKKRKGKN-NCKQVSEGDINDEVDHBDLADGFKFDLADDTT 81  
QY 70 ---GWEI--SGFEKKGAGVYINISKOFFIPVFK-----IEKKKEENRPTD 114  
DB 82 SNFGQWNLARSGSNNDAAFAV--KQVDLTKITRKGGLYMAHIDSKQEEETKE-K 138  
QY 115 VSKKKDN-----PQVNSQLNSHREKEDLQREHSOKSDSTKQVT 154  
DB 139 VKENDSDDEBLAMDGFQKADPANNQDENQSEEEBEEKEEKEEERBEQBQEBT 192

RESULT 34  
DRL1 YEAST STANDARD; PRT; 1007 AA.  
AC P39083; P39934;  
DT 01-FEB-1995 (Rel. 31, Created)  
DT 01-FEB-1995 (Rel. 31, Last sequence update)  
DT 15-SEP-2003 (Rel. 42, Last annotation update)  
DE Rho-type GTPase-activating protein 1.  
GN RGA1 OR DBM1 OR TRH1 OR YOR127W OR O3290 OR YOR3290W.  
OS Saccharomyces cerevisiae (Baker's Yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.  
OX NCBI\_TaxID=4932;  
RN (1)  
RN SEQUENCE FROM N.A., AND MUTAGENESIS.  
RP STRAIN=S288C;  
RX MEDLINE=96239492; PubMed=8657111;  
RA Chen G.-C., Zheng L., Chan C.S.M.;  
RT "The LIM domain-containing Dbm1 GTPase-activating protein is required  
RT for normal cellular morphogenesis in Saccharomyces cerevisiae.";  
RT Mol. Cell. Biol. 16:1376-1390 (1996).  
RL (2)  
RN SEQUENCE FROM N.A.  
RP STRAIN=S288C / FY1679;  
RX MEDLINE=97060020; PubMed=8904341;  
RA Wiemann S., Reckmann S., Benes V., Voss H., Schwager C., Visek C.,  
RA Stegmann J., Zimmermann J., Brille H., Paces V., Ansgore W.;  
RT "Sequencing and analysis of 51 kb on the right arm of chromosome XV  
RT from Saccharomyces cerevisiae reveals 30 open reading frames.";  
RL Yeast 12:281-288 (1996).  
RN (3)  
RN SEQUENCE FROM N.A.  
RP MEDLINE=97344368; PubMed=9200815;  
RX Voss H., Benes V., Andrade M.A., Valencia A., Reckmann S., Teodoru C.,  
RA Schwager C., Paces V., Sander C., Ansgore W.;  
RT "DNA sequencing and analysis of 130 kb from yeast chromosome XV.";  
RL Yeast 13:655-672 (1997).  
RN (4)  
RN SEQUENCE FROM N.A.  
RP STRAIN=S288C;  
RX MEDLINE=96101594; PubMed=7498791;  
RA Stevenson B.J., Ferguson B., de Virgilio C., Bi E., Pringle J.R.,

RA	Ammerer G., Sprague G.F., Jr.:		
RA	"Mutation of RGA1, which encodes a putative GTPase-activating protein		
RT	for the polarity-establishment protein Cdc42p, activates the		
RT	phenomene-response pathway in the yeast <i>Saccharomyces cerevisiae</i> ."		
RL	Genes Dev. 9:2949-2963 (1995).		
RN	[5]		
RP	SEQUENCE OF 570-639 FROM N.A.		
RC	STRAIN-SNY243.		
RX	MEDLINE=93087574; PubMed=1454852;		
RA	Ramer S.W., Elledge S.J., Davis R.W.:		
RT	"Dominant genetics using a yeast genomic library under the control of		
RT	a strong inducible promoter."		
RL	Proc. Natl. Acad. Sci. U.S.A. 89:11589-11593 (1992).		
CC	-1- FUNCTION: GTPASE-ACTIVATING PROTEIN (GAP) FOR CDC42 AND/OR RHO1.		
CC	NEGATIVE REGULATOR OF THE PHENOMENE-RESPONSE PATHWAY THROUGH THE		
CC	STE20 PROTEIN KINASE; ACTS AT A STEP BETWEEN THE G-PROTEIN AND THE		
CC	MAP KINASE MODULE. DOMINANT SUPPRESSOR OF BUD EMERGENCE DEFECT		
CC	CAUSED BY DELETION OF IPL2/BEM2. INVOLVED IN THE CONTROL OF		
CC	POLARIZED CELL GROWTH AND PROPER BUD SITE SELECTION.		
CC	-1- SIMILARITY: Contains 2 LIM zinc-binding domains.		
CC	-1- SIMILARITY: Contains 1 Rho-GAP domain.		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (see <a href="http://www.isb-sib.ch/announce/or_send_an_email_to_license@isb-sib.ch">http://www.isb-sib.ch/announce/</a>		
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).		
CC	-----		
DR	EMBL; U07421; AAA16875.1; -.		
DR	EMBL; X90518; CAA62108.1; -.		
DR	EMBL; X94335; CAA64046.1; -.		
DR	EMBL; Z75035; CAA9326.1; -.		
DR	EMBL; X90950; CAA62445.1; -.		
DR	EMBL; L02617; AAA35153.1; -.		
DR	PIR; S48535; S48535.		
DR	SGD; S0005653; RGA1.		
DR	GO; GO:0005187; P:Rho GTPase activator activity; IPI.		
DR	GO; GO:0004871; P:signal transducer activity; IPI.		
DR	GO; GO:0007015; P:actin filament organization; IMP.		
DR	GO; GO:0007118; P:apical bud growth; IPI.		
DR	GO; GO:0000283; P:establishment of cell polarity (sensu Sacch. .); IPI.		
DR	GO; GO:0007125; P:Invasive growth; IPI.		
DR	GO; GO:0007119; P:isotropic bud growth; IPI.		
DR	GO; GO:0007120; P:pseudohyphal growth; IPI.		
DR	GO; GO:0000755; P:signal transduction during conjugation with. .; IGI.		
DR	GO; GO:0007264; P:small GTPase mediated signal transduction; IPI.		
DR	InterPro; IPR001181; LIM.		
DR	InterPro; IPR000198; RhoGAP.		
DR	Pfam; PF00412; LIM; 1.		
DR	Pfam; PF00620; RhoGAP; 1.		
DR	ProDom; PD000094; LIM; 2.		
DR	SMART; SM00132; LIM; 2.		
DR	SMART; SM00134; RhoGAP; 1.		
DR	PROSITE; PS00478; LIM DOMAIN_1; 1.		
DR	PROSITE; PS50023; LIM DOMAIN_2; 2.		
DR	PROSITE; PS50238; RHOGAP; 1.		
KM	GTPase activation; Repeat; LIM domain; Metal-binding; Zinc;		
KM	Phenomene response.		
FT	DOMAIN 13 66 LIM 1.		
FT	DOMAIN 70 122 LIM 2.		
FT	DOMAIN 791 1006 RHO-GAP.		
FT	VARIANT 866 866 V -> A.		
FT	VARIANT 898 898 K -> R.		
FT	VARIANT 926 926 S -> G.		
FT	MUTAGEN 40 40 C->S; BIPOLAR BUDDING.		
FT	MUTAGEN 37 37 C->S; BIPOLAR BUDDING.		
FT	MUTAGEN 98 98 C->S; BIPOLAR BUDDING.		
FT	MUTAGEN 101 101 C->S; BIPOLAR BUDDING.		
FT	CONFLICT 457 457 T -> E (IN REF. 4).		
FT	CONFLICT 507 507 D -> P (IN REF. 4).		
GO	SEQUENCE 1007 AA; 112831 MF; CB05411B57553791 CRC64;		

Query Match 9.5%; Score 85.5; DB 1; Length 1007;  
Best Local Similarity 21.0%; Pred. No. 78;  
Matches 34; Conservative 34; Mismatches 69; Indels 25; Gaps 3;

DY DTGVESELKPRVTVTITQNGEKMSSTI-----VSEDFILPVYKGELKNGYQFDGEIS 74  
::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|  
Db ITTGISIQMDSLSKRVANNIVEETNALQRPVVEVKEDRSVPDLAIGVQEGAEKYTSNN 408  
349 NGSKKDAGGVINLSKDTFIKPVFVKIEEKEBENKTFPDVSXKKGNPQVNHSOLNESH 134  
DY GFGEKKDAGGVINLSKDTFIKPVFVKIEEKEBENKTFPDVSXKKGNPQVNHSOLNESH 134  
::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|  
Db 409 SGGRKRITSRLSRSSKOLMI-----NLKSRAFGKDSNVKLSPAKVTISR 454  
135 KEDIQREHSHQSDDSYKVATATVD-----KNNISASTNN 171  
DY ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|  
Db 455 SODLRNDSDHTGTDPNSNSTLIDLVNNOKSINYRFTDN 496

RESULT 35

MAPP\_RAT STANDARD; PRT; 2459 AA.

ID MAPB\_RAT  
AC P15205; Q6Z958; Q9ER21; Q9QW92;  
DT 01-APR-1990 (Rel. 14, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Microtubule-associated protein 1B (MAP 1B) (Neuraxin) [contains: MAP1 light chain LC1].  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
CX NCBI\_faxid=10116;  
RN [1]  
RP SEQUENCE OF 1-142 FROM N.A.  
RC STRAIN=Sprague-Dawley; TISSUE=Testis;  
RX MEDLINB=36257242; PubMed=8665295;  
RA Liu D., Fischer I.;  
RT "Isolation and sequencing of the 5' end of the rat microtubule-associated protein (MAP1B)-encoding cDNA";  
RT Gene 172:307-308(1996).  
RL

RN [2]  
RP SEQUENCE OF 96-2459 FROM N.A., DOMAIN, AND INDUCTION.  
RC STRAIN=Sprague-Dawley; TISSUE=Brain, and Gli3 tumor;  
RX MEDLINB=92347374; PubMed=1639092;  
RA Zauner W., Kratz J., Staunton J., Feick P., Wiche G.;  
RT "Identification of two distinct microtubule binding domains on recombinant rat MAP 1B";  
RT Eur. J. Cell Biol. 57:66-74(1992).  
RL

RN [3]  
RP SEQUENCE OF 1541-2459 FROM N.A., AND TISSUE SPECIFICITY.  
RC TISSUE=Spinal cord;  
RX MEDLINB=90059871; PubMed=2555150;  
RA Rientz A., Grenningloh G., Hermans-Borgmeyer I., Kirsch J., Littauer U.Z., Prior P., Gundelfinger E.D., Schmitt B., Betz H.;  
RT "Neuraxin, a novel putative structural protein of the rat central nervous system that is immunologically related to microtubule-associated protein 5.";  
RT EMBO J. 8:2879-2888(1989).  
RL

[4]  
RN DEVELOPMENTAL STAGE, AND PHOSPHORYLATION.  
RP MEDLINB=97405699; PubMed=2260743;  
RX Ma D., Notthias F., Byrne L.J., Fischer I.;  
RT "Differential regulation of microtubule-associated protein 1B (MAP1B) in rat CNS and PNS during development";  
RT J. Neurosci. Res. 49:319-332(1997).  
RL

-1- FUNCTION: The function of brain MAPs is essentially unknown. Phosphorylated MAP1B may play a role in the cytoskeletal changes that accompany neurite extension. Possibly MAP1B binds to at least two tubulin subunits in the polymer, and this bridging of subunits might be involved in nucleating microtubule polymerization and in stabilizing microtubules.  
CC -1-SUBUNIT: 3 different light chains, LC1, LC2 and LC3, can associate with MAP1A and MAP1B proteins.  
CC

```

RESULT 36
PENN_MESOR STANDARD; PRT; 365 AA.
ID PENN_MESOR
AC P41343;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Ferredoxin-NADP reductase, chloroplast precursor (EC 1.18.1.2) (FNR).
GN PERH OR FNRA.
OS Mesembryanthemum crystallinum (Common ice plant).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Caryophyllales; Caryophyllales; Aizoaceae; Mesembryanthemum.
OC NCBI_TextID=3544;
RN [1]
RP SEQUENCE FROM N.A.
RA Michalowski C.B., Schmitt J.M., Bohnert H.J.;
RT "Expression during salt stress and nucleotide sequence of cDNA for
RT ferredoxin-NADP+ reductase from Mesembryanthemum crystallinum.";
RL Plant Physiol. 89:817-822(1989).
CC -1- FUNCTION: MAY PLAY A KEY ROLE IN REGULATING THE RELATIVE AMOUNTS
CC OF CYCLIC AND NON-CYCLIC ELECTRON FLOW TO MEET THE DEMANDS OF THE
CC PLANT FOR ATP AND REDUCING POWER.
CC -1- CATALYTIC ACTIVITY: Reduced ferredoxin + NADP(+) = oxidized
CC ferredoxin + NADPH.
CC -1- COFACTOR: PAD.
CC -1- PATHWAY: FINAL STEP IN LINEAR PHOTOSYNTHETIC ELECTRON TRANSPORT
CC CHAIN, IT HAS ALSO BEEN IMPLICATED IN CYCLIC ELECTRON FLOW AROUND
CC PHOTOSYSTEM I, ITS ROLE BEING TO RETURN ELECTRONS FROM FERREDOXIN
CC TO THE CYTOCHROME B-P COMPLEX.
CC -1- SUBCELLULAR LOCATION: CHLOROPLAST; STROMAL SIDE OF THE THYLAKOID
CC MEMBRANE IN THE VICINITY OF THE PHOTOSYSTEM I IN THE NON-STACKED
CC AND FRINSE PORTION OF THE MEMBRANE.
CC -1- SIMILARITY: WITH OTHER SPECIES FNR.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as their content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/announcement/
CC or send an email to license@ebi.ac.uk).
-----
DR EMBL; M25528; AAA33029.1; -.
DR PIR; A44974; A44974.
DR HSPD; P00455; IFNC.
DR InterPro; IPR001709; PFN_cyt_reductse.
DR InterPro; IPR001433; OxRed_FAD/NAD(P)+.
DR Pfam; PF00175; NAD_Binding_1; 1.
DR PRINTS; PR00371; FENCRC.
KW Oxidoreductase; Flavoprotein; NADP; FAD; Chloroplast; Translact peptide;
KW Electron transport; Photosynthesis; Thylakoid; Membrane.
FT TRANSIT ? CHLOROPLAST (POSTTRANSIT)
FT GRAIN ? 365 FERREDOXIN-NADP REDUCTASE.
FT DISULFID 183 188 BY SIMILARITY.
FT NP BIND 217 235 NADP (RIBOSE PART) (BY SIMILARITY).
SQ SEQUENCE 365 AA; 41063 MW; 02A10BF6E6C15FC CRC64;
Query Match 9.5%; Score 85; DB 1; Length 365;
Best Local Similarity 28.1%; Pred. No. 28;
Matches 41; Conservative 20; Mismatches 45; Indels 40; Gaps 9;
QY DEPARNT---VKEFLKTKGTGEV-----SELKP-HRVVTYIQTNGKEM-----SS 45
DB DFGDSKTVALCVKRLITNDNGELVKVCNSPLCDLKPSBVLUTGVGSGEMLPKOPNA 214
155
QY TIV-SHEDELTP---VKGLELEK-GYOPDG--WEISGEKKAGAVININSKOTFIK 95
DB TIMLANGTGIAIPRSPLTMGMPEFKDDYENGLAMLFGLG-----VPTSSLLYK 264
96 PVFKKIEKKBEENKPPVDVSKKDN 121

```

DB 265 EPEKMEKAPENRDLPAVSRQTN 290

RESULT 37

6PGD LACLA STANDARD: PRT: 472 AA.

AC 09CHU6;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE 6-phosphogluconate dehydrogenase (EC 1.1.1.44).

GN GND OR L10622.

OS Lactococcus lactis (subsp. lactis) (Streptococcus lactis).

CC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Lactococcus.

CC NCBI\_TaxId=1360;

CC [1]

CC SEQUENCE FROM N.A.

CC STRAIN=IL1403;

CC MEDLINE=21235186; PubMed=11337471;

CC RA Bolotin A., Wincker P., Manger S., Jallion O., Malarme K.,

CC RA Weisenbach J., Ehrlich S.D., Sorokin A.,

CC RT "The complete genome sequence of the lactic acid bacterium Lactococcus

CC RT lactis ssp. lactis IL1403."

CC RL Genome Res. 11:731-753(2001).

CC CC -1- CATALYTIC ACTIVITY: 6-phospho-D-gluconate + NADP(+) = D-ribulose

CC CC 5-phosphate + CO(2) + NADPH.

CC CC -1- PATHWAY: Hexose monophosphate shunt.

CC CC -1- SIMILARITY: BELONGS TO THE 6-PHOSPHOGLUCONATE DEHYDROGENASE

CC CC FAMILY.

CC CC -----

CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC CC the European Bioinformatics Institute. There are no restrictions on its

CC CC use by non-profit institutions as long as its content is in no way

CC CC modified and this statement is not removed. Usage by and for commercial

CC CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).

CC CC -----

CC CC EMBL; AE006295; AAK04720.1; -

CC CC PIR; F86702; F86702.

CC CC HSSP; P00349; 2PGD.

CC CC InterPro; IPR006183; 6PGD.

CC CC InterPro; IPR006114; 6PGD C.

CC CC InterPro; IPR006113; 6PGD decarbox.

CC CC InterPro; IPR006115; 6PGD\_NAD.

CC CC InterPro; IPR006184; 6PGDdm.

CC CC Pfam; PF00393; 6PGD; 1.

CC CC Pfam; PF03446; NAD binding 2; 1.

CC CC PRINTS; PR00076; 6PGDHGNSB.

CC CC TIGRPFAMs; TIGR00873; gnd; 1.

CC CC PROSITE; PS00461; 6PGD; 1.

CC CC DR Gluconate utilization; Oxidoreductase; Pentose shunt; NADP;

CC CC KW Complete proteome.

CC CC SEQUENCE 472 AA; 52424 MW; 6EF4F504F217E55F CRC64;

Query Match 9.5%; Score 85; DB 1; Length 472;

Best Local Similarity 22.9%; Pred. No. 37;

Matches 32; Conservative 23; Mismatches 33; Indels 52; Gaps 7;

QY 5 KDPARNTVKEPIKNGTGEVSELKPHRVVTYIQNGKMSSTIYSEEPILPVY-KGEL- 62

DB 49 KNLVLTITLIERFV-----GSLE--KPRIMLMVQGAATDAITKS-----LPLLDKGDIL 97

QY 63 -----EKGYOPDGEISGF-----GKKDGYVINLSKD 91

DB 98 IDGNTHTPPPTMRNNAELADSGINFTIGVSGEGKALLGPSMMPGGQKEA-----Y 149

QY 92 TPIKPVFKIEKKEENKP 111

DB 150 DLVAPIPEQIAAKAPQDGKP 169

RESULT 38

Y040\_MYCN

ID Y040\_MYCN STANDARD: PRT: 657 AA.

AC P75062;

DT 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Hypothetical lipoprotein MG040 homolog precursor (D09\_orf557).

GN MPO52 OR MPO102.

OS Mycoplasma pneumoniae.

CC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.

CC NCBI\_TaxId=2104;

CC [1]

CC SEQUENCE FROM N.A.

CC STRAIN=ATCC 29342 / M129;

CC MEDLINE=97105885; PubMed=8948633;

CC RA Himelreich R., Hilbert H., Plagens H., Pirkl E., Li B.-C.,

CC RA Hermann R.;

CC RT "Complete sequence analysis of the genome of the bacterium Mycoplasma

CC RT pneumoniae."

CC RL Nucleic Acids Res. 24:4420-4449(1996).

CC CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a lipid anchor

CC CC (potential).

CC CC -1- SIMILARITY: SOME, TO T. PALLIDIUM TMPC.

CC CC -----

CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC CC the European Bioinformatics Institute. There are no restrictions on its

CC CC use by non-profit institutions as long as its content is in no way

CC CC modified and this statement is not removed. Usage by and for commercial

CC CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).

CC CC -----

CC CC EMBL; AE000012; AAB95750.1; -

CC CC PIR; S73428; S73428.

CC CC InterPro; IPR003760; Bmp.

CC CC Pfam; PF02608; Bmp; 1.

CC CC PROSITE; PS00013; PROKAR\_LIPOPROTEIN; 1.

CC CC DR Hypothetical protein; Lipoprotein; Membrane; Signal;

CC CC KW Complete proteome.

CC CC FT SIGNAL 1 26 POTENTIAL.

CC CC FT CHAIN 27 657 HYPOTHEICAL LIPOPROTEIN MG040 HOMOLOG.

CC CC FT LIPID 27 27 N-ACTYL DIGLYCERIDE (POTENTIAL).

CC CC SEQUENCE 657 AA; 16711 MW; 6147B1A5673606A9 CRC64;

Query Match 9.5%; Score 85; DB 1; Length 657;

Best Local Similarity 16.7%; Pred. No. 53;

Matches 46; Conservative 38; Mismatches 81; Indels 110; Gaps 6;

QY 9 RNTTVKEPIKND-----TGEVSELKPHRVVTYIQNGKEM----- 43

DB 319 RATTINNLDKGVSAIIPVAGPQVNLAVNEVARKKATTAIGVDSAGELDIINDAPDK 378

QY 44 -----SSTIVSEBDFILPVYGBLEKGYOPDG----- 70

DB 379 DQLIKKNKIIIPSSIKALDVALEMLAIOKGSNNYKGGVNNIGVTSSVGISBA 438

QY 71 -----WEISGFEKQDAGVINI-----SKDTPIKPVFKIEKKE----- 106

DB 439 GYEFLLIDPVFWKTQSGSKAMATMTNLKRLSSDPTNKKALKVSTHNGSDXGIIICK 498

QY 107 -----EENKPTPVY-SKKQDNPOVNSQLAESHRKEDL 138

DB 499 YSNLLTKKSTVTAVAQKSMTDNNSGTQKQKLSVDITKCKEKSAGKGTSGNSGDGQON 558

QY 139 QREHSQKSDSTKDTATVADKNNISSTNNPN 173

DB 559 GKETNDIISKYSKLLTMTTNNKVNSSKKQSSDN 593

RESULT 39

YALE\_SCHPO STANDARD: PRT: 719 AA.

ID YALE\_SCHPO

AC Q10342; Q9URU8;

```

DT 01-OCT-1996 (Rel. 34, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DZ Hypothetical protein C106.14c in chromosome 11.
CN SPEC106.14c.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomyces.
OX NCBI_TaxId=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=21848401; PubMed=11899360;
RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgourou J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Felwell T., Frazer A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holtroyd S., Hornby T., Howarth S., Huckle E.J., Hunt S., Jagsels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odeh C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch B.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volckaert G., Aert R., Robben J., Grynolprez B.,
RA Woljens I., Volckaert E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hillbert H.,
RA Borzym K., Langer I., Beck A., Lehnach H., Reinhardt R., Pohl T.M.,
RA Beger P., Zimmermann W., Medler H., Wandut R., Furnelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA Galbert P., Aves S., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet G., Gallardin C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forzburg S.L.,
RA Cernutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Shpakovski G.V., Useery D., Barrett B.G., Nurse P.;
RA "The genome sequence of Schizosaccharomyces pombe."
RT Nature 415:871-880(2002).
RL -1- SIMILARITY: TO YEAST YGR245C.
CC -1-
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL, AL110295; CAB5730.1; -.
CC DR PIR, T39271; T39271.
CC DR Genedb.SPombe; SPEC106.14c; -.
CC DR Pfam; PF05285; SDAL; 1.
CC KW Hypothetical protein.
CC FT DOMAIN 250 254 POLY-GLU.
CC FT DOMAIN 559 562 POLY-GLU.
CC SQ SEQUENCE 719 AA; 81250 MW; 938645586BD1D6 CRC64;

Query Match 9.5%; Score 85; DB 1; Length 719;
Best Local Similarity 23.9%; Pred. No. 59;
Matches 34; Conservative 32; Mismatches 62; Indels 14; Gaps 5;

QY 15 EPILNKDTGEVSELPKRVATVITONGKMSSTVSEEDFILPVYKGL-----EKGYQFPG 70
DB 469 DMLKKRDKRKLASIEKDKRT-PLKGEELANTHGIGLELQYKAKSHBEGNGDMDN 527
QY 71 WEIS-GFBEKQDAGVIVINSKDTFKIPVKIKIBKKKEENKPTFDVSKKQNDPOVNHSQL 129
DB 528 WEVSEDDGQNSDDSGGIDVDSDNI-----ELSDSDEBEKAT---ARKSEDEKSSGQK 579
QY 130 NESHRKEDLOREBHSOKSDSTK 151

```

```

DB 580 ELVDKMTSLASQSLTPNDLKR 601

RESULT 40
ID IF2_STAEP STANDARD; PRT; 720 AA.
AC Q8CST4;
DT 15-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Translation initiation factor IF-2.
GN INFB OR S80945.
OS Staphylococcus epidermidis.
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxId=1282;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 12228;
RA Zhang Y., Ren S., Li H., Fu G., Lu L., Lu G., Jia J., Tu Y., Qin Z.,
RA Chen Z., Wen Y.;
RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: One of the essential components for the initiation of
CC protein synthesis. Protects formylmethionyl-tRNA from spontaneous
CC hydrolysis and promotes its binding to the 30S ribosomal subunits.
CC Also involved in the hydrolysis of GTP during the formation of the
CC 70S ribosomal complex (by similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: BELONGS TO THE IF-2 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL, AE016747; AA004542.1; -.
CC DR HAMAP; MF 00100; -.
CC DR Pfam; PF06009; GTP_EPTU; 1.
CC DR Pfam; PF03144; GTP_BPTU_D2; 1.
CC DR Pfam; PF04760; IF2_N; 2.
CC DR Prodom; PD186100; IF2_N; 2.
CC DR TIGRPFAM; TIGR00487; IF-2; 1.
CC DR TIGRPFAM; TIGR00231; small_GTP; 1.
CC DR PROSITE; PS01176; IF2; 1.
CC KW Initiation factor; Protein biosynthesis; GTP-binding;
CC KW Complete proteome.
CC FT DOMAIN 225 373 G-DOMAIN.
CC FT NP_BIND 231 238 GTP (BY SIMILARITY).
CC FT NP_BIND 277 281 GTP (BY SIMILARITY).
CC FT NP_BIND 331 334 GTP (BY SIMILARITY).
CC SQ SEQUENCE 720 AA; 79343 MW; 07FB5A6A59CF970C CRC64;

Query Match 9.5%; Score 85; DB 1; Length 720;
Best Local Similarity 28.0%; Pred. No. 59;
Matches 23; Conservative 15; Mismatches 36; Indels 8; Gaps 2;

QY 101 IEKKEEENKPTFDVSKKQD-----NPQVHSQLNESHRKEDLOREBHSOKSDSTK-----D 152
DB 38 LEEBQIKALDKFKPKASQAKDTYKQNTQNHQSKNSKNDKQSKNSKKTCKKQEN 97
QY 153 VTATVLDKNNISKSXTNNPNK 174
DB 98 NKGQKQKNNKNTKNNQKNNK 119

```

Search completed: February 10, 2004, 10:49:56  
Job time : 9.08902 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 10, 2004, 10:48:44 ; Search time 37.3096 Seconds  
(without alignments)  
697.707 Million cell updates/sec

Title: US-10-067-385-8\_COPY\_610\_773  
Perfect score: 848  
Sequence: 1 TTYKERTLNKDTGSEVSLKP.....ATVLAKNNISSKSTTNPNK 164

Scoring table: BL0SUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_19Jun03.\*  
1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*  
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.\*  
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	848	100.0	773	22	AA848343
2	848	100.0	2120	21	AA581710
3	848	100.0	2140	24	ABU01020
4	615	72.5	117	19	AAW5096
5	615	72.5	117	23	ABP54590
6	119	14.0	746	22	AA681779
7	119	14.0	778	23	ABP39023
8	107.5	12.7	484	21	AA647777
9	106	12.5	665	21	AA818278

10	101.5	12.0	564	22	ABB61977
11	99.5	11.7	1384	24	ABP55413
12	99.5	11.7	1404	24	ABP55393
13	97.5	11.5	2519	22	ABG16636
14	96.5	11.4	408	21	AA371134
15	96.5	11.4	456	21	AA371133
16	96.5	11.4	476	21	AA371132
17	96.5	11.4	2060	23	AAE20967
18	96	11.3	2500	21	AA818272
19	93.5	11.0	645	24	ABJ19106
20	93.5	11.0	654	24	ABP56879
21	92.5	10.9	645	24	ABJ18979
22	92	10.8	299	22	AA682397
23	92	10.8	309	23	ABP40822
24	91	10.7	442	23	ABP38188
25	91	10.7	902	23	ABP40312
26	90.5	10.7	209	21	AA842897
27	90.5	10.7	635	23	ABP73209
28	90.5	10.7	757	21	AA44364
29	90.5	10.7	758	21	AA853319
30	90.5	10.7	801	23	ABP68963
31	90.5	10.7	922	22	AA685023
32	90	10.6	1183	22	ABP58769
33	89.5	10.6	402	20	AA403190
34	89.5	10.6	511	20	AA435091
35	89.5	10.6	516	23	ABP73570
36	89.5	10.6	525	20	AA843189
37	89.5	10.6	1408	22	ABP58704
38	89	10.5	296	23	ABP93341
39	88.5	10.4	281	22	ABP64828
40	88.5	10.4	1141	22	AA685008
41	88	10.4	225	23	ABP73992
42	88	10.4	258	22	AA84584
43	88	10.4	817	22	AAW79318
44	88	10.4	817	22	AAW79319
45	88	10.4	2515	22	ABP12281

#### ALIGNMENTS

RESULT 1  
AA848343  
ID AA848343 standard; Protein: 773 AA.  
AC AA848343;  
XX  
XX  
DT 20-APR-2001 (first entry)  
XX  
XX  
DE S. pneumoniae Spi30 polypeptide.  
XX  
XX  
KW Immunogenic; Spi28; Spi30; pneumococcal; otitis media; nasopharyngeal;  
KW bronchial; lung; blood; infection; immune response; immunotherapy;  
KW antibacterial; auditory; vaccine.  
XX  
XX  
OS Streptococcus pneumoniae.  
XX  
XX  
PN WO200076540-A2.  
XX  
XX  
PD 21-DEC-2000.  
XX  
XX  
PF 09-JUN-2000; 2000WO-US15925.  
XX  
XX  
PR 10-JUN-1999; 99US-0138453.  
XX  
XX  
PA (MED) MED IMMUNE INC.  
XX  
XX  
PI Adamou JE, Choi GH;  
XX  
XX  
DR WPI, 2001-112197/12.  
DR N-PsDB; AAC84742.  
XX  
XX  
PT New vaccines comprising Spi28 or Spi30 polypeptides, for treating and

Drosophila melanog  
Human MDR-22 prot  
Human MDR-2 prote  
Novel human diagno  
Arabidopsis thalia  
Arabidopsis thalia  
Staphylococcus lug  
Plasmodium falcipa  
Pathogen specific  
Staphylococcus epi  
Pathogen specific  
S. epidermidis ope  
Staphylococcus epi  
Staphylococcus epi  
Human ORFX ORP261  
Human albicans e  
Human cell cycle r  
Human colon cancer  
Human polypeptide  
Shrimp white spot  
Drosophila melanog  
S. aureus trigger  
Candida albicans e  
S. aureus trigger  
Drosophila melanog  
Herbicidally activ  
Drosophila melanog  
Shrimp white spot  
Candida albicans e  
Human protein sequ  
Human protein SEQ  
Human secreted pro



PT or ear infection -  
XX  
PS Claim 1; SEQ ID No 1180; 56pp; English.  
XX  
XX The invention relates to a protein comprising or having at least 508  
CC identity to any of the 2469 amino acid sequences, identified in the  
CC specification (available on a computer readable format), or its fragment,  
CC expressed from 2469 of 2489 identified DNA coding regions from the  
CC Streptococcus pneumoniae type 4 strain genomic sequence appearing as  
CC AB556454. Also included are an antibody which binds one of the  
CC proteins, treating a patient by administering the protein, DNA or  
CC antibody (in a composition), a kit comprising first and second primers,  
CC which are the nucleic acid cited above or fragments between nucleotides  
CC 8-100 of a sequence not defined in the specification, for amplifying a  
CC target sequence contained within a Streptococcus nucleic acid sequence,  
CC where the first primer is substantially complementary to the target  
CC sequence and the second primer is substantially complementary to the  
CC complement of the target sequence, and where the parts of the primers  
CC having substantial complementarity define the termini of the target  
CC sequence to be amplified, assay comprising contacting a test compound  
CC with the protein, and determining whether the test compound binds to the  
CC protein and a Streptococcus pneumoniae bacterium, where one or more  
CC genes encoding the proteins has been rendered inactive. The proteins,  
CC nucleic acid molecules, antibody and compositions are useful as  
CC medicaments for treating or preventing a disease or infection due to  
CC streptococcus bacteria, particularly S. pneumoniae, such as pneumonia,  
CC sepsis, otitis media or ear infection. They are also useful in developing  
CC vaccines, diagnostics and antibiotics. The methods are useful for  
CC identifying immunodominant proteins. The present sequence is one of  
CC the 2469 proteins expressed by the identified coding regions from the  
CC genomic sequence.  
CC Note: The sequence data for this patent did not form part  
CC of the printed specification, but was obtained in electronic  
CC format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 2140 AA;  
Query Match 100.0%; Score 848; DB 24; Length 2140;  
Best Local Similarity 100.0%; Pred. No. 4.4e-74;  
Matches 164; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 TTVEEPLINKDGEVSEIKPHRYVTITQNGKMSSTIYSEEDFILPYKGELEKGYPDG 60  
DB 1943 TTVEEPLINKDGEVSEIKPHRYVTITQNGKMSSTIYSEEDFILPYKGELEKGYPDG 2002  
QY 61 WEISFEGSKDAGYINLSKDTPIKVPFKIIEKKEENKPTFDVSKKKDNPQVNSQLN 120  
DB 2003 WEISFEGSKDAGYINLSKDTPIKVPFKIIEKKEENKPTFDVSKKKDNPQVNSQLN 2062  
QY 121 ESHKEDLQREHSOKSDSTKQVTAIVLDKNNISSKSTTNPNK 164  
DB 2063 ESHKEDLQREHSOKSDSTKQVTAIVLDKNNISSKSTTNPNK 2106

RESULT 4  
AAW55096  
ID AAW55096 standard; Protein; 117 AA.  
XX  
AC AAW55096;  
XX  
DT 02-OCT-1998 (first entry)  
XX  
DE Streptococcus pneumoniae SP043 protein.  
XX  
KM Streptococcus pneumoniae; antigen; vaccine; infection; diagnosis;  
XX detection; pneumonia; otitis media; meningitis.  
XX  
OS Streptococcus pneumoniae.  
XX  
PN M09818930-A2.  
XX  
PD 07-MAY-1998.

XX  
PF 30-OCT-1997; 97WO-US19422.  
XX  
PR 31-OCT-1996; 96US-0029960.  
XX  
PA (HUMA-) HUMAN GENOME SCI INC.  
XX  
PI Choi GH, Hromockyj A, Johnson LS, Kunsch CA;  
XX WPI; 1998-272224/24.  
DR N-PSDB; AAW27357.  
DR  
XX  
PT Nucleic acid encoding antigenic peptide(s) from Streptococcus  
XX pneumoniae - or their epitope-containing fragments, useful in  
XX protective or therapeutic vaccines, and for diagnosis  
XX  
PS Claim 11; Page 62; 118pp; English.  
XX  
CC The present sequence represents a protein from Streptococcus pneumoniae.  
CC The nucleic acid sequence encoding the Streptococcus pneumoniae protein  
CC can be useful in vaccines for inducing protective antibodies against  
CC Streptococcus pneumoniae, for treatment or prevention of infection e.g.  
CC pneumonia, otitis media or meningitis. Probes based on the nucleic acid  
CC are used to detect Streptococcus infection (by usual hybridization or  
CC amplification methods), also for isolating Streptococcus genes or their  
CC allelic variants. The protein can be used similarly to detect specific  
CC antibodies in standard immunoassays, especially for diagnosing or  
CC monitoring infections. Antibodies which bind the protein are used to  
CC detect corresponding antigens, to purify the protein and for passive  
CC immunisation (optionally coupled to a toxin). Vaccines are administered,  
CC e.g. by injection, orally or through the skin, typically at 0.01-1000  
CC (especially 10-300) mu g/ml per dose.  
XX  
SQ Sequence 117 AA;  
Query Match 72.5%; Score 615; DB 19; Length 117;  
Best Local Similarity 100.0%; Pred. No. 7.3e-53;  
Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 48 YKGELEKGYQPDGWEISFEGSKDAGYINLSKDTPIKVPFKIIEKKEENKPTFDVSK 107  
DB 1 YKGELEKGYQPDGWEISFEGSKDAGYINLSKDTPIKVPFKIIEKKEENKPTFDVSK 60  
QY 108 KKDNPQVNSQLNESHKEDLQREHSOKSDSTKQVTAIVLDKNNISSKSTTNPNK 164  
DB 61 KKDNPQVNSQLNESHKEDLQREHSOKSDSTKQVTAIVLDKNNISSKSTTNPNK 117

RESULT 5  
ABP54590  
ID ABP54590 standard; Protein; 117 AA.  
XX  
AC ABP54590;  
XX  
DT 04-SEP-2002 (first entry)  
XX  
DE S. pneumoniae SP043 protein sequence SEQ ID NO:68.  
XX  
KM Streptococcus pneumoniae; epitope; vaccine; antigenic protein;  
XX antibacterial; Streptococcal infection; detection.  
XX  
OS Streptococcus pneumoniae.  
XX  
PN US2002061545-A1.  
XX  
PD 23-MAY-2002.  
XX  
PF 22-JAN-2001; 2001US-0765272.  
XX  
PR 30-OCT-1997; 97US-0961083.  
XX  
PA (CHOI/) CHOI G H.  
XX (KUNSC/) KUNSC C A.

PA (BARA/) BARASH S C.  
PA (DILL/) DILLON P J.  
PA (DOUG/) DOUGHERTY B.  
PA (FANN/) FANNON M R.  
PA (ROSE/) ROSEN C A.  
XX  
PI Choi GH, Kunech CA, Barash SC, Dillon PJ, Dougherty B, Fannon MR,  
PI Rosen CA;  
XX  
DR WPI; 2002-479261/51.  
DR N-PSDB; ABQ84825.  
XX  
PT New Streptococcus pneumoniae antigens, useful for detecting  
PT Streptococcus and for preventing or attenuating disease caused by  
PT Streptococcus infection -  
XX  
PS Claim 11; Page 29; 70pp; English.  
XX  
CC ABQ84792 to ABQ84904 represents nucleic acids which encode the  
CC Streptococcus pneumoniae antigens given in ABP54557 to ABP54669.  
CC The S. pneumoniae antigens have antibacterial activity and can be  
CC used in vaccines. The S. pneumoniae antigens can also be used to  
CC prevent or attenuate a Streptococcal infection in an animal. The  
CC polynucleotides encoding the S. pneumoniae antigens can be used to  
CC detect Streptococcus nucleic acids. ABQ84905 to ABQ85130 represent  
CC primers used in the cloning of S. pneumoniae ORFs (open reading frames)  
CC which are used in an example from the present invention.  
XX  
SQ Sequence 117 AA;  
XX  
Query Match 72.5%; Score 615; DB 23; Length 117;  
Best Local Similarity 100.0%; Pred. No. 7.3e-53;  
Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
XX  
QY 48 YNGELEKGYQPDGMBISGFEKGDAGYVNLKDTFIKPVFKKIEKKKEENKPTDVS 107  
DB 1 YNGELEKGYQPDGMBISGFEKGDAGYVNLKDTFIKPVFKKIEKKKEENKPTDVS 60  
XX  
QY 108 KKDNPQVNSQNLNESHKEDLQREHSQKSDSTKDVATVLDNNISKSSTNNPNK 164  
DB 61 KKDNPQVNSQNLNESHKEDLQREHSQKSDSTKDVATVLDNNISKSSTNNPNK 117  
XX  
RESULT 6  
AAG81779  
ID AAG81779 standard; Protein; 746 AA.  
XX  
AC AAG81779;  
XX  
DT 03-SEP-2001 (first entry)  
XX  
DE S. epidermidis open reading frame protein sequence SEQ ID NO:652.  
XX  
KM Staphylococcus epidermidis SRI strain; infection; diagnosis;  
XX  
KW vaccination; endocarditis.  
XX  
OS Staphylococcus epidermidis.  
XX  
PN MO200134809-A2.  
XX  
PD 17-MAY-2001.  
XX  
PF 09-NOV-2000; 2000NO-US030782.  
XX  
PR 09-NOV-1999; 99US-0164258.  
XX  
PA (GLAX) GLAXO GROUP LTD.  
XX  
PI Kimerly MJ;  
XX  
DR WPI; 2001-316495/33.  
DR N-PSDB; AAH52629.  
XX

PT Nucleic acids encoding polypeptides from Staphylococcus epidermidis,  
PT useful for vaccinating against infections, e.g. endocarditis -  
XX  
PS Claim 18; Page 208; 218pp; English.  
XX  
CC AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides  
CC (II), given in AAG81454 to AAG81120, from Staphylococcus epidermidis.  
CC (I) and (II) can have antibacterial activity and therefore can be used  
CC in vaccination. The nucleic acids (I) may be used to produce the  
CC S. epidermidis polypeptides (II) via the production of vectors  
CC containing them which are used to produce hosts cells which express the  
CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be  
CC used to vaccinate subjects and to raise antibodies against the bacteria.  
CC The polypeptides may also be used to assay for other inhibitors of their  
CC activity and therefore identify compounds that may be used for the  
CC treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to  
CC AAH55090 represent specifically claimed S. epidermidis genomic DNA  
CC polynucleotide sequences from the present invention. AAH55091 to  
CC AAH55098 represent oligonucleotide sequences and primers which are used  
CC in the exemplification of the present invention.  
CC N.B. The present invention specifically claims all the polynucleotide  
CC sequences given in the sequence listing of the present specification,  
CC however the sequence listing only goes up to SEQ ID NO:4454 so even  
CC though sequences are given in the disclosure for SEQ ID NO:4465 to 4472,  
CC no sequences are present for SEQ ID NO:4455 to 4464.  
XX  
SQ Sequence 746 AA;  
XX  
Query Match 14.0%; Score 119; DB 22; Length 746;  
Best Local Similarity 27.0%; Pred. No. 0.0081;  
Matches 47; Conservative 24; Mismatches 57; Indels 46; Gaps 9;  
XX  
QY 10 KNGEVELKPHVATVTLONGKEMSTVISEBDPFIIPVK-----GELEKGYQPDGM-- 61  
DB 585 EDSVNAQSLKP-----ITLNGKQIKQOQSVKSTKVLPHSKVLMATDGLTMP-DMTGWT 639  
XX  
QY 62 -EISGFE-----GKKDAGYVIN--LSKDTFIKPVFKKIEKKKEENKPTDVS---K 107  
DB 640 EDTLAFEDLTLYKSTKNGCFVFNOSTSKQITK-----NCKIEVSLSAED 686  
XX  
QY 108 KKDNPQVNSQNLNESHKEDLQREHSQKSDSTKDVATVLDNNISKSSTNN 161  
DB 687 TDDQEKTDDESDSNKSKDKADBDHSNTSSSTKN-----DGSNMDSKNDSD 734  
XX  
RESULT 7  
ABP39023  
ID ABP39023 standard; Protein; 778 AA.  
XX  
AC ABP39023;  
XX  
DT 24-JUL-2002 (first entry)  
XX  
DE Staphylococcus epidermidis ORF amino acid sequence SEQ ID NO:3868.  
XX  
KM Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;  
XX  
KW antibacterial; gene therapy.  
XX  
OS Staphylococcus epidermidis.  
XX  
PN US6380370-B1.  
XX  
PD 30-APR-2002.  
XX  
PF 13-AUG-1998; 98US-0134001.  
XX  
PR 14-AUG-1997; 97US-055779P.  
XX  
PR 08-NOV-1997; 97US-064964P.  
XX  
PA (GENO-) GENCOMB THERAPEUTICS CORP.  
XX  
PI Doucette-Stamm LA, Bush D;  
XX





PT Plasmodium falciparum, useful as antimalarial vaccines and in the  
PT diagnosis of P.falciparum infection -  
XX  
XX  
PS Disclosure; Page 321-322; 577pp; English.  
XX  
CC The present invention describes proteins and their fragments (I) encoded  
CC by chromosome 2 of the human malarial parasite, Plasmodium falciparum.  
CC Also described are: (1) nucleotide sequences (II) encoding (I); and (2)  
CC vaccines against P. falciparum infection comprising (I) or (II).  
CC (1) and (II) are useful for the development of vaccines against  
CC P. falciparum infection. (1) and polyclonal antisera or a monoclonal  
CC antibody raised to immunogens comprising the sequences of (1), are  
CC useful in the detection of infection with P. falciparum. Furthermore,  
CC (1) (especially when they are rifins or secreted or membrane proteins)  
CC can aid the identification of drugs to treat or prevent P. falciparum  
CC infection, or they can be used to identify drug resistance in  
CC P. falciparum. Sequencing of the Plasmodium chromosome 2 and the  
CC subsequent identification of proteins encoded by it will help to expand  
CC our understanding of parasite biology, a process hampered by the  
CC complexity of the parasitic life cycle, and provide new targets for  
CC vaccine and drug development. Parasite resistance to drugs and mosquito  
CC resistance to insecticides have led to a resurgence of malaria in many  
CC parts of the world, and there is a pressing need for vaccines and new  
CC drugs. AAA70078 to AAA70287 and AAB18144 to AAB18352 represent nucleotide  
CC and protein sequences given in the present invention, but which are not  
CC specifically mentioned within the specification.  
XX  
SQ Sequence 665 AA;  
Query Match 12.5%; Score 106; DB 21; Length 665;  
Best Local Similarity 24.3%; Pred. No. 0.13; Indels 56; Gaps 10;  
Matches 42; Conservative 35; Mismatches 40;  
QY 10 KDTGEVSLKPRRT-VTIQNGKEMSTVSEEDFILPVYKGLKGYQPDGWEI--SGF 66  
DB 127 EKKKIKKSDLRKQNEALNLSGK-----NEQDI-----NKKKKGQ-----DISNNA 169  
QY 67 EKKKAGYVINTSKDTPIKPVFKKIEKKK-----BNKPTPD-----VSKKKDNP 112  
DB 170 ENKKD-----VKEGVKELEKKKKKKKESKSDHKVBNKKSDHKVBNKKSDDH 217  
QY 113 QVNHSQLNSHHRKDLOR-BEHSQKSGSDTKDVTATVLDKNNISSTKTTNNPNK 164  
DB 218 KVENKKKSDHKIEBKVKVKEHEDEDEB-----DKKKKSENKKKDKENK 261  
RESULT 10  
ABB61977  
ID ABB61977 standard; Protein; 564 AA.  
AC ABB61977;  
XX  
XX 26-MAR-2002 (first entry)  
XX  
DE Drosophila melanogaster polypeptide SEQ ID NO 12723.  
XX  
XX Drosophila; developmental biology; cell signalling; insecticide;  
XX pharmaceutical.  
XX  
XX Drosophila melanogaster.  
XX  
XX WO200171042-A2.  
XX  
XX 27-SEP-2001.  
XX  
XX 23-MAR-2001; 2001WO-US09231.  
XX  
XX 23-MAR-2000; 2000US-191637P.  
XX 11-JUL-2000; 2000US-0614150.  
XX  
XX (PEKE ) PE CORP NY.  
XX  
XX Venter JC, Adams M, Li PWD, Myers EW;  
PI

XX  
XX WPI; 2001-656860/75.  
DR N-PSDB; ABL06080.  
XX  
XX  
PT New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions -  
XX  
XX  
PS Disclosure; SEQ ID NO 12723; 21pp + Sequence listing; English.  
XX  
CC The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA  
CC sequences (AB101840-AB16175) and the encoded proteins  
CC (AB157737-AB12072).  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 564 AA;  
Query Match 12.0%; Score 101.5; DB 22; Length 564;  
Best Local Similarity 24.5%; Pred. No. 0.29;  
Matches 34; Conservative 29; Mismatches 51; Indels 25; Gaps 5;  
QY 41 EDFILPVYKGLKGYQPDGWEI-----EISGFEKKAGYVI-----NLKDTPIK 85  
DB 78 EDLDTPILESNFSK--VFDGVDHEDHEDHGDVPEGBALDDHEDHDDHEDDBBE 135  
QY 86 PVFKKIEKKKEENKPT-----PDVSKKKNPQVNHSQLNSHHRKDLORBSHSQKSDS 139  
DB 136 PLTELEBELREBERPEDEBPADVEYEBDEENNA--GENTLAEDAEEREBEEDMD 193  
QY 140 TKDVTATVLDKNNISSTKST 158  
DB 194 EGTVEATVETATTEATTEAT 212  
RESULT 11  
ABP55413  
ID ABP55413 standard; protein; 1384 AA.  
XX  
XX ABP55413;  
XX  
XX 04-FEB-2003 (first entry)  
XX  
XX Human MDDT-22 protein SEQ ID NO:22.  
XX  
XX  
XX Human; MDDT; molecules for disease detection and treatment; anti-HIV;  
XX antiallergic; antiinflammatory; antianemic; antiparkinsonian; nootropic;  
XX immunosuppressive; antifertility; antiretroviral; antidiabetic;  
XX antidiabetic; nephrotoxic; antitumor; cytostatic; hepatotropic; dermatological;  
XX osteopathic; antirheumatic; antiparasitic; antihelminthic; antiparasitic;  
XX uropathic; ophthalmological; antirheumatic; haemostatic; antibacterial;  
XX virucide; protozoacide; fungicide; gene therapy; cell proliferative;  
XX cancer; developmental disorder; neurological disorder; infection;  
XX reproductive disorder; autoimmune disorder; inflammatory disorder.  
XX  
XX Homo sapiens.  
XX  
XX WO200278420-A2.  
XX  
XX 10-OCT-2002.  
XX  
XX 29-MAR-2002; 2002WO-US09809.  
XX  
XX 30-MAR-2001; 2001US-280387P.  
XX 05-APR-2001; 2001US-282335P.  
XX 13-APR-2001; 2001US-283663P.  
XX

PR 19-APR-2001; 2001US-285484P.  
 PR 18-JAN-2002; 2002US-350702P.  
 PR 25-JAN-2002; 2002US-351749P.  
 XX  
 XX  
 PA (INCY-) INCYTE GENOMICS INC.  
 PI Lu DAM, Arvizu CS, Gandhi AR, Hafalia AJA, Ding L, Lu Y;  
 PI Rantumar J, Swarnakar A, Tang YT, Yue H, Tran B, Lee SY;  
 PI Warren BA, Nguyen DB, Thangavelu K, Yao MG, Elliott VS, Baughn MR;  
 PI Emerling BW, Lai PG, Gietzen KJ, Becha SD, Marquis JP, Kable AE;  
 PI MPI; 2003-058385/05.  
 DR N-PSDB; ABQ83880.  
 XX  
 PT New human molecules for disease detection and treatment, useful for  
 PT diagnosing, treating or preventing autoimmune or inflammatory disorders  
 PT (e.g. AIDS, allergy or anemia), multiple sclerosis, osteoarthritis,  
 PT cancer or hepatitis -  
 XX  
 PS Claim 1; Page 207-211; 238pp; English.  
 XX  
 XX The present invention describes 23 human molecules for disease detection  
 CC and treatment (MDDT-1 to 23) (see ABP55392 to ABP55414). The human  
 CC MDDT-1 to 23 proteins (I) are encoded by the sequences given in ABQ83859  
 CC to ABQ83881. (I) can have various activities depending on the cells and  
 CC tissues in which they are expressed. These activities include: anti-HIV;  
 CC antiallergic; antiinflammatory; antianaemic; antiparkinsonian; nootropic;  
 CC anticonvulsant; antinfertility; antiarteriosclerotic; antiasclerotic;  
 CC immunosuppressive; antithyroid; cyrostatic; hepatotropic; dermatological;  
 CC antidiabetic; nephrotropic; antigout; thyromimetic; neuroprotective;  
 CC osteopathic; ophthalmological; antiparasitic; antihelminthic; antibacterial;  
 CC uropathic; protozoacide; and fungicide. (I) and the polynucleotides  
 CC encoding them can be used in gene therapy. (I), polynucleotides, agonists  
 CC and antagonists from the present invention can be used for diagnosing,  
 CC treating or preventing disorders associated with aberrant expression of  
 CC MDDT, particularly cell proliferative (e.g. cancer), developmental  
 CC disorders, neurological disorders, reproductive disorders, or  
 CC autoimmune/inflammatory disorders, or viral, bacterial, fungal,  
 CC parasitic, protozoal or helminthic infections. They are also useful in  
 CC the assessment of the effects of exogenous compounds on the expression  
 CC of nucleic acid and amino acid sequences of proteins associated with  
 CC MDDT.  
 XX  
 XX Sequence 1384 AA;  
 SQ  
 Query Match 11.7%; Score 99.5; DB 24; Length 1384;  
 Best Local Similarity 21.7%; Pred. No. 1.6;  
 Matches 45; Conservative 32; Mismatches 67; Indels 63; Gaps 7;  
 QY 4 KEPIANKDTGVSSEKLPKRVVTYTONGKMSSTVSEBDFILPVYKGLKGYQF----- 58  
 DB 1013 KEPIMLQNGEISOLK-KEIERIQMGKMSVWKQEOYIATQYKALDLCQELRLTRE 1071  
 QY 59 -----DQWELSGRGKDA-----GVYINLSK----- 80  
 DB 1072 QVONSHTELAARHQVQAOREIERLSLIEBKOLSKKDKMAGNLAELGASKYREAH 1131  
 QY 81 -----DTPIKVP/KKIEBKKE-----BENKFPEDVSKKKDNQVNSHQSLSHRE 126  
 DB 1132 LEARQAQAEIKKLSAEVESLLEAVNEMISHOEHNAMKIS--ADSQKSSVQQLNEQLERA 1189  
 QY 127 DLQREHSQSDSTKDVATVLDKNKI 153  
 DB 1190 KLELEB---ADPTVSNLHQGVODRNEV 1213  
 RESULT 12  
 ABP55393  
 ID ABP55393 standard; protein; 1404 AA.  
 XX  
 AC ABP55393;  
 XX

DT 04-FEB-2003 (first entry)  
 XX  
 XX Human MDDT-2 protein SEQ ID NO:2.  
 DE  
 XX  
 XX Human; MDDT; molecules for disease detection and treatment; anti-HIV;  
 KW antiallergic; antiinflammatory; antianaemic; antiparkinsonian; nootropic;  
 KW immunosuppressive; antithyroid; cyrostatic; hepatotropic; dermatological;  
 KW antidiabetic; nephrotropic; antigout; thyromimetic; neuroprotective;  
 KW osteopathic; antiparasitic; antihelminthic; antibacterial;  
 KW uropathic; ophthalmological; antiparasitic; haemostatic; antiparasitic;  
 KW vrusicide; protozoacide; fungicide; gene therapy; cell proliferative;  
 KW cancer; developmental disorder; neurological disorder; infection;  
 KW reproductive disorder; autoimmune disorder; inflammatory disorder.  
 XX  
 OS Homo sapiens.  
 XX  
 XX MO200278420-A2.  
 PN  
 XX  
 PD 10-OCT-2002.  
 XX  
 XX 29-MAR-2002; 2002WO-US09809.  
 PF  
 XX  
 XX 30-MAR-2001; 2001US-280387P.  
 PR 05-APR-2001; 2001US-282335P.  
 PR 13-APR-2001; 2001US-283663P.  
 PR 19-APR-2001; 2001US-285484P.  
 PR 18-JAN-2002; 2002US-350702P.  
 PR 25-JAN-2002; 2002US-351749P.  
 XX  
 XX (INCY-) INCYTE GENOMICS INC.  
 PA  
 XX  
 XX Lu DAM, Arvizu CS, Gandhi AR, Hafalia AJA, Ding L, Lu Y;  
 PI Rantumar J, Swarnakar A, Tang YT, Yue H, Tran B, Lee SY;  
 PI Warren BA, Nguyen DB, Thangavelu K, Yao MG, Elliott VS, Baughn MR;  
 PI Emerling BW, Lai PG, Gietzen KJ, Becha SD, Marquis JP, Kable AE;  
 PI MPI; 2003-058385/05.  
 DR N-PSDB; ABQ83860.  
 XX  
 PT New human molecules for disease detection and treatment, useful for  
 PT diagnosing, treating or preventing autoimmune or inflammatory disorders  
 PT (e.g. AIDS, allergy or anemia), multiple sclerosis, osteoarthritis,  
 PT cancer or hepatitis -  
 XX  
 XX  
 PS Claim 1; Page 172-175; 238pp; English.  
 XX  
 XX The present invention describes 23 human molecules for disease detection  
 CC and treatment (MDDT-1 to 23) (see ABP55392 to ABP55414). The human  
 CC MDDT-1 to 23 proteins (I) are encoded by the sequences given in ABQ83859  
 CC to ABQ83881. (I) can have various activities depending on the cells and  
 CC tissues in which they are expressed. These activities include: anti-HIV;  
 CC antiallergic; antiinflammatory; antianaemic; antiparkinsonian; nootropic;  
 CC anticonvulsant; antinfertility; antiarteriosclerotic; antiasclerotic;  
 CC immunosuppressive; antithyroid; cyrostatic; hepatotropic; dermatological;  
 CC antidiabetic; nephrotropic; antigout; thyromimetic; neuroprotective;  
 CC osteopathic; ophthalmological; antiparasitic; antihelminthic; antibacterial;  
 CC uropathic; protozoacide; and fungicide. (I) and the polynucleotides  
 CC encoding them can be used in gene therapy. (I), polynucleotides, agonists  
 CC and antagonists from the present invention can be used for diagnosing,  
 CC treating or preventing disorders associated with aberrant expression of  
 CC MDDT, particularly cell proliferative (e.g. cancer), developmental  
 CC disorders, neurological disorders, reproductive disorders, or  
 CC autoimmune/inflammatory disorders, or viral, bacterial, fungal,  
 CC parasitic, protozoal or helminthic infections. They are also useful in  
 CC the assessment of the effects of exogenous compounds on the expression  
 CC of nucleic acid and amino acid sequences of proteins associated with  
 CC MDDT.  
 XX  
 XX Sequence 1404 AA;  
 SQ  
 Query Match 11.7%; Score 99.5; DB 24; Length 1404;  
 XX



Best Local Similarity 21.7%, Pred. No. 1.6,  
Matches 45; Conservative 32; Mismatches 67; Indels 63; Gaps 7;

```
OY      4 KEFLIKGTGSELPKPRVVTIQTNGMGSESTIVSEDPILPYKGLBKGYQF----- 58  
Db     1033 KEFLIMQNEQELISQL-KRIERTQQRMKEMSVMKEQIQATQYKAIDLJGQELRTRE 1091  
OY      59 -----DGEISGFEEKDA-----GVINLSK----- 80  
Db     1092 QVONSHTLEARHQVOQAQRIERIELTSELSEBEMKMLSKCAKHGNHLAEFGASGRVRAH 1151  
OY      81 -----DTPIKVPFKKIEKKR-----EENKETFPVSKKQDNPPVNSQLNESHKE 126  
Db     1152 LEARMQAKIKKLGAVESLKEAYIHMEWISHQENNAKMRLS--ADSQKSSVQQLNQLEKA 1209  
OY      127 DLQREHSQKSDSTQDVATATVLDKNIT 153  
Db     1210 KLELE---AODTVSNLHQVQDDRENV 1233
```

RESULT 13

ABG1636  
ID ABG1636 standard; Protein; 2519 AA.  
XX  
AC ABG1636;  
DT 18-FEB-2002 (first entry)  
XX  
DE Novel human diagnostic protein #16627.  
XX  
HM Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KM food supplement; medical imaging; diagnostic; genetic disorder.  
XX  
OS Homo sapiens.  
XX  
WO200175067-A2.  
PN  
PD 11-OCT-2001.  
XX  
PF 30-MAR-2001; 2001WO-US08631.  
XX  
PR 31-MAR-2000; 2000US-0540217.  
PR 23-AUG-2000; 2000US-0649167.  
XX  
PA (HYSB-) HYSEQ INC.  
XX  
P1 Drmanac RT, Liu C, Tang YT;  
DR MPI; 2001-639362/73.  
XX  
DR N-PSDB; AAS80823.

New isolated polypeptide and encoded polypeptides, useful in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits and to assess biodiversity -

Claim 20; SEQ ID No 46995; 103bp; English.

The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity

[illegible]

PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 17-JUN-1999; 99US-0139452.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.

PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0146389.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147263.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.

PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161921.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 11.4%; Score 96.5; DB 21; Length 408;  
Best Local Similarity 25.0%; Pred. No. 0.58;  
Matches 40; Conservative 29; Mismatches 60; Indels 31; Gaps 8;

QY 10 KDTGEVSEIKP-HVVTYTIQNGKSSSTIVGEEDFILVYNGELEGYQPDGWEISGREG 68  
DB 197 KNKKEKQPLKPKHPVSAFLVYANERRALREKNSVAVAK-----ITGEEKRNLS 248  
QY 69 KKDAGY--VINLSKDTFKPVFKLIEKKEEENKPTFPVSKKKQNPQVNHQSLNESHKRE 126  
DB 249 KKAAPYKAKVAKNKETIYQ-AMEYKRTKEE-----ALSQKE-----EEELKLNKOE 297  
QY 127 DLGREHSOKSDSTKDVATATVLDKNNISSKSTNN--PNK 164  
DB 298 ALQWLKKEKKTND-----LIKKEKATKKKKNNENVDPNK 330

RESULT 15  
AAG37133  
ID AAG37133 standard; Protein; 456 AA.

AC AAG37133;  
DT 18-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 45609.

KM Protein identification; signal transduction pathway; metabolic pathway;  
KM hybridisation assay; genetic mapping; gene expression control; promoter;  
KM termination sequence.

OS Arabidopsis thaliana.

XX BP1033405-A2.

PN 06-SEP-2000.

XX 25-FEB-2000; 2000EP-0301439.

PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130419.  
PR 23-APR-1999; 99US-0130510.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 07-MAY-1999; 99US-0132487.  
PR 11-MAY-1999; 99US-0132863.  
PR 14-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.

PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144684.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145215.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.

	PR	26-JUL-1999;	99US-0145276.
	FR	27-JUL-1999;	99US-0145913.
	PR	28-JUL-1999;	99US-0145918.
	PR	28-JUL-1999;	99US-0145919.
	PR	02-AUG-1999;	99US-0146386.
	PR	02-AUG-1999;	99US-0146388.
	PR	02-AUG-1999;	99US-0146389.
	PR	03-AUG-1999;	99US-0147038.
	PR	04-AUG-1999;	99US-0147204.
	PR	04-AUG-1999;	99US-0147302.
	PR	05-AUG-1999;	99US-0147192.
	PR	05-AUG-1999;	99US-0147260.
	PR	06-AUG-1999;	99US-0147303.
	PR	06-AUG-1999;	99US-0147416.
	PR	09-AUG-1999;	99US-0147493.
	PR	09-AUG-1999;	99US-0147935.
	PR	10-AUG-1999;	99US-0148171.
	PR	11-AUG-1999;	99US-0148319.
	PR	12-AUG-1999;	99US-0148341.
	PR	13-AUG-1999;	99US-0148565.
	PR	13-AUG-1999;	99US-0148684.
	PR	16-AUG-1999;	99US-0149368.
	PR	17-AUG-1999;	99US-0149175.
	PR	18-AUG-1999;	99US-0149426.
	PR	20-AUG-1999;	99US-0149722.
	PR	20-AUG-1999;	99US-0149723.
	PR	20-AUG-1999;	99US-0149929.
	PR	23-AUG-1999;	99US-0149902.
	PR	23-AUG-1999;	99US-0149930.
	PR	25-AUG-1999;	99US-0150566.
	PR	26-AUG-1999;	99US-0150884.
	PR	27-AUG-1999;	99US-0151065.
	PR	27-AUG-1999;	99US-0151066.
	PR	27-AUG-1999;	99US-0151080.
	PR	30-AUG-1999;	99US-0151303.
	PR	31-AUG-1999;	99US-0151438.
	PR	01-SEP-1999;	99US-0151930.
	PR	07-SEP-1999;	99US-0152363.
	PR	10-SEP-1999;	99US-0153070.
	PR	13-SEP-1999;	99US-0153758.
	PR	15-SEP-1999;	99US-0154018.
	PR	16-SEP-1999;	99US-0154033.
	PR	20-SEP-1999;	99US-0154779.
	PR	22-SEP-1999;	99US-0155139.
	PR	23-SEP-1999;	99US-0155486.
	PR	24-SEP-1999;	99US-0155659.
	PR	28-SEP-1999;	99US-0156458.
	PR	29-SEP-1999;	99US-0156596.
	PR	04-OCT-1999;	99US-0157117.
	PR	05-OCT-1999;	99US-0157753.
	PR	06-OCT-1999;	99US-0157865.
	PR	07-OCT-1999;	99US-0158029.
	PR	08-OCT-1999;	99US-0158232.
	PR	12-OCT-1999;	99US-0158369.
	PR	13-OCT-1999;	99US-0159293.
	PR	13-OCT-1999;	99US-0159294.
	PR	13-OCT-1999;	99US-0159299.
	PR	14-OCT-1999;	99US-0159329.
	PR	14-OCT-1999;	99US-0159329.
	PR	14-OCT-1999;	99US-0159330.
	PR	14-OCT-1999;	99US-0159331.
	PR	14-OCT-1999;	99US-0159637.
	PR	14-OCT-1999;	99US-0159638.
	PR	18-OCT-1999;	99US-0159584.
	PR	21-OCT-1999;	99US-0160741.
	PR	21-OCT-1999;	99US-0160767.
	PR	21-OCT-1999;	99US-0160768.
	PR	21-OCT-1999;	99US-0160770.
	PR	21-OCT-1999;	99US-0160814.
	PR	21-OCT-1999;	99US-0160815.
	PR	22-OCT-1999;	99US-0160980.
	PR	22-OCT-1999;	99US-0160981.
	PR	22-OCT-1999;	99US-0160989.
	PR	26-JUL-1999;	99US-0145276.
	FR	27-JUL-1999;	99US-0145913.
	PR	28-JUL-1999;	99US-0145918.
	PR	28-JUL-1999;	99US-0145919.
	PR	02-AUG-1999;	99US-0146386.
	PR	02-AUG-1999;	99US-0146388.
	PR	02-AUG-1999;	99US-0146389.
	PR	03-AUG-1999;	99US-0147038.
	PR	04-AUG-1999;	99US-0147204.
	PR	04-AUG-1999;	99US-0147302.
	PR	05-AUG-1999;	99US-0147192.
	PR	05-AUG-1999;	99US-0147260.
	PR	06-AUG-1999;	99US-0147303.
	PR	06-AUG-1999;	99US-0147416.
	PR	09-AUG-1999;	99US-0147493.
	PR	09-AUG-1999;	99US-0147935.
	PR	10-AUG-1999;	99US-0148171.
	PR	11-AUG-1999;	99US-0148319.
	PR	12-AUG-1999;	99US-0148341.
	PR	13-AUG-1999;	99US-0148565.

PR 14-MAY-1999;	99US-0134218.
PR 14-MAY-1999;	99US-0134219.
PR 14-MAY-1999;	99US-0134221.
PR 14-MAY-1999;	99US-0134370.
PR 18-MAY-1999;	99US-0134768.
PR 19-MAY-1999;	99US-0134941.
PR 20-MAY-1999;	99US-0135124.
PR 21-MAY-1999;	99US-0135353.
PR 24-MAY-1999;	99US-0135629.
PR 25-MAY-1999;	99US-0136021.
PR 27-MAY-1999;	99US-0136392.
PR 28-MAY-1999;	99US-0136782.
PR 01-JUN-1999;	99US-0137222.
PR 03-JUN-1999;	99US-0137528.
PR 04-JUN-1999;	99US-0137502.
PR 07-JUN-1999;	99US-0137724.
PR 08-JUN-1999;	99US-0138094.
PR 10-JUN-1999;	99US-0138540.
PR 10-JUN-1999;	99US-0138847.
PR 14-JUN-1999;	99US-0139119.
PR 16-JUN-1999;	99US-0139452.
PR 17-JUN-1999;	99US-0139452.
PR 18-JUN-1999;	99US-0139454.
PR 18-JUN-1999;	99US-0139455.
PR 18-JUN-1999;	99US-0139456.
PR 18-JUN-1999;	99US-0139457.
PR 18-JUN-1999;	99US-0139458.
PR 18-JUN-1999;	99US-0139459.
PR 18-JUN-1999;	99US-0139460.
PR 18-JUN-1999;	99US-0139461.
PR 18-JUN-1999;	99US-0139462.
PR 18-JUN-1999;	99US-0139463.
PR 18-JUN-1999;	99US-0139750.
PR 21-JUN-1999;	99US-0139817.
PR 22-JUN-1999;	99US-0139889.
PR 23-JUN-1999;	99US-0140353.
PR 24-JUN-1999;	99US-0140354.
PR 24-JUN-1999;	99US-0140695.
PR 28-JUN-1999;	99US-0140823.
PR 29-JUN-1999;	99US-0140991.
PR 30-JUN-1999;	99US-0141287.
PR 01-JUL-1999;	99US-0141842.
PR 01-JUL-1999;	99US-0142154.
PR 02-JUL-1999;	99US-0142055.
PR 06-JUL-1999;	99US-0142390.
PR 08-JUL-1999;	99US-0142803.
PR 09-JUL-1999;	99US-0142920.
PR 12-JUL-1999;	99US-0142977.
PR 13-JUL-1999;	99US-0143542.
PR 14-JUL-1999;	99US-0143624.
PR 15-JUL-1999;	99US-0144005.
PR 16-JUL-1999;	99US-0144086.
PR 19-JUL-1999;	99US-0144332.
PR 19-JUL-1999;	99US-0144332.
PR 19-JUL-1999;	99US-0144333.
PR 19-JUL-1999;	99US-0144334.
PR 19-JUL-1999;	99US-0144335.
PR 20-JUL-1999;	99US-0144352.
PR 20-JUL-1999;	99US-0144632.
PR 20-JUL-1999;	99US-0144884.
PR 21-JUL-1999;	99US-0144814.
PR 21-JUL-1999;	99US-0145086.
PR 21-JUL-1999;	99US-0145088.
PR 22-JUL-1999;	99US-0145085.
PR 22-JUL-1999;	99US-0145087.
PR 22-JUL-1999;	99US-0145089.
PR 22-JUL-1999;	99US-0145195.
PR 23-JUL-1999;	99US-0145218.
PR 23-JUL-1999;	99US-0145218.
PR 23-JUL-1999;	99US-0145224.
PR 27-JUL-1999;	99US-0145913.
PR 27-JUL-1999;	99US-0145918.
PR 27-JUL-1999;	99US-0145919.
PR 28-JUL-1999;	99US-0145951.
PR 02-AUG-1999;	99US-0146386.
PR 02-AUG-1999;	99US-0146388.
PR 03-AUG-1999;	99US-0146389.
PR 03-AUG-1999;	99US-0147038.
PR 04-AUG-1999;	99US-0147204.
PR 04-AUG-1999;	99US-0147302.
PR 05-AUG-1999;	99US-0147192.
PR 05-AUG-1999;	99US-0147303.
PR 06-AUG-1999;	99US-0147303.
PR 06-AUG-1999;	99US-0147416.
PR 09-AUG-1999;	99US-0147493.
PR 09-AUG-1999;	99US-0147935.
PR 10-AUG-1999;	99US-0148171.
PR 11-AUG-1999;	99US-0148319.
PR 12-AUG-1999;	99US-0148341.
PR 13-AUG-1999;	99US-0148565.
PR 13-AUG-1999;	99US-0148684.
PR 16-AUG-1999;	99US-0149368.
PR 17-AUG-1999;	99US-0149175.
PR 18-AUG-1999;	99US-0149426.
PR 20-AUG-1999;	99US-0149722.
PR 20-AUG-1999;	99US-0149723.
PR 20-AUG-1999;	99US-0149929.
PR 23-AUG-1999;	99US-0149902.
PR 23-AUG-1999;	99US-0149930.
PR 25-AUG-1999;	99US-0150566.
PR 26-AUG-1999;	99US-0150884.
PR 27-AUG-1999;	99US-0151065.
PR 27-AUG-1999;	99US-0151066.
PR 27-AUG-1999;	99US-0151080.
PR 30-AUG-1999;	99US-0151303.
PR 31-AUG-1999;	99US-0151438.
PR 01-SEP-1999;	99US-0151930.
PR 07-SEP-1999;	99US-0152363.
PR 10-SEP-1999;	99US-0153070.
PR 13-SEP-1999;	99US-0153758.
PR 15-SEP-1999;	99US-0154018.
PR 16-SEP-1999;	99US-0154039.
PR 20-SEP-1999;	99US-0154779.
PR 22-SEP-1999;	99US-0155139.
PR 23-SEP-1999;	99US-0155486.
PR 24-SEP-1999;	99US-0155659.
PR 28-SEP-1999;	99US-0156458.
PR 29-SEP-1999;	99US-0156596.
PR 04-OCT-1999;	99US-0157117.
PR 05-OCT-1999;	99US-0157753.
PR 06-OCT-1999;	99US-0157865.
PR 07-OCT-1999;	99US-0158029.
PR 08-OCT-1999;	99US-0158232.
PR 12-OCT-1999;	99US-0158369.
PR 13-OCT-1999;	99US-0159293.
PR 13-OCT-1999;	99US-0159294.
PR 13-OCT-1999;	99US-0159329.
PR 14-OCT-1999;	99US-0159329.
PR 14-OCT-1999;	99US-0159330.
PR 14-OCT-1999;	99US-0159331.
PR 14-OCT-1999;	99US-0159637.
PR 18-OCT-1999;	99US-0159638.
PR 18-OCT-1999;	99US-0159584.
PR 21-OCT-1999;	99US-0160741.
PR 21-OCT-1999;	99US-0160767.
PR 21-OCT-1999;	99US-0160768.
PR 21-OCT-1999;	99US-0160770.
PR 21-OCT-1999;	99US-0160814.
PR 21-OCT-1999;	99US-0160815.
PR 22-OCT-1999;	99US-0160980.
PR 22-OCT-1999;	99US-0160981.

```
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match
Best Local Similarity 25.0%; Score 96.5; DB 21; Length 476;
Matches 40; Conservative 29; Mismatches 60; Indels 31; Gaps 8;

QY 10 KDTGEVSEIKP-HRVTTIIONGKMSSTIVSEDFLLPYKGELEKGYFDGWEISGFEFG 68
DB 265 KKKKKKDKLPKHPVSAFLVYANERPAALREKKSVYEVAK-----ITGGEWNLSD 316
QY 69 KDDAGT--VINISKDTPIKPVFKIIEKKKEENKPTFDVSKKKDNQVNHSQLNESHRE 126
DB 317 KKKAPYEKVKKKKKEVYLQ-AMEEYKRTKEE-----ALSQKKE-----EEELKLKQK 365
QY 127 DQREHNSQSDSTKDTATVLDKNNISKSTNN--PMK 164
DB 366 ALQMLKKKKEKTDN-----LKKERATKKKKNNENVDPMK 398

RESULT 17
AAE20967
ID AAE20967 standard; Protein; 2060 AA.
XX
XX AAE20967;
XX
XX 01-JUL-2002 (first entry)
XX
DE Staphylococcus lugdunensis von Willebrand factor binding protein.
XX
XX Von Willebrand factor binding protein; vwb; immunogen; antibacterial;
XX
XX vaccine; infection.
XX
XX Staphylococcus lugdunensis.
XX
XX
XX Key Location/Qualifiers
XX FT 1..47
XX FT /label= Signal_peptide
XX FT 48..2060
XX FT /note= "Mature von Willebrand factor binding protein"
XX
XX MO200228892-A1.
XX
XX 11-APR-2002.
XX
XX 06-APR-2001; 2001WO-SE00766.
XX
XX 04-OCT-2000; 2000SE-0003573.
XX
XX (BIOS-) BIOSTAPRO AB.
XX
XX Guss B, Frykberg L, Jacobsson K, Ahlen J, Nilsson M;
XX
XX WPI; 2002-304928/34.
XX
XX N-PSDB; AAD33371.
XX
XX
XX New von Willebrand factor binding protein from Staphylococci, useful
XX
XX for determining and treating staphylococcal infection -
XX
XX
XX Claim 3; Page 35-41; 53pp; English.
XX
XX The present invention relates to von Willebrand factor binding protein or
XX
XX polypeptide (vwb) from Staphylococci. The vwb and immunogens of vwb are
XX
XX useful in vaccines to combat infections caused by Staphylococci. The
```

```
CC invention is also useful for detection of staphylococcal infection and
CC purifying von Willebrand factor from a complex solution. The present
CC sequence is Staphylococcus lugdunensis vwb1 protein.
XX
XX
SQ Sequence 2060 AA;
XX
Query Match
Best Local Similarity 25.7%; Score 96.5; DB 23; Length 2060;
Matches 39; Conservative 25; Mismatches 49; Indels 39; Gaps 9;

QY 14 EVSEIKPVRVVTIIO-NGKMSSTIVSEDFLLPYKGELEKGYFDG--WEISGFEK 70
DB 1914 DIDEIK---ITIVDTNGRE-----IVPSRKGLP-EGFIGDMQYTGK--- 1955
QY 71 DAGVINISKDTPIKPVFKIIE-----KKEENKPTFDVSKKKDNQVNHSQLNESH 123
DB 1956 -----IEKQITITTYKVENAVPAQLKKTEN--TQSSQFGHPTQVAVQQLVKYAN 2006
QY 124 RRE--DQREHNSQSDSTKDTATVLDKNNI 153
DB 2007 VKEQSIKSEHTDMVSELPTGETANNGL 2038

RESULT 18
AAB18272
ID AAB18272 standard; Protein; 2500 AA.
XX
XX AAB18272;
XX
XX 07-NOV-2000 (first entry)
XX
DE Plasmodium falciparum chromosome 2 related protein SEQ ID NO:129.
XX
XX Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;
XX
XX antimalarial; malaria; protozoacide; infection; insecticide.
XX
XX Plasmodium falciparum.
XX
XX MO200025728-A2.
XX
XX 11-MAY-2000.
XX
XX 05-NOV-1999; 99WO-US26796.
XX
XX 05-NOV-1998; 98US-0107131.
XX
XX (HOFF/) HOFFMAN S.
XX PA (CARD/) CARUCCI D.
XX PA (GARD/) GARDNER M.
XX PA (VENT/) VENTER J C.
XX
XX Hoffman S, Carucci D, Gardner M, Venter JC;
XX
XX WPI; 2000-365347/31.
XX
XX
XX Proteins encoded by chromosome 2 of the human malarial parasite,
XX
XX Plasmodium falciparum, useful as antimalarial vaccines and in the
XX
XX diagnosis of P. falciparum infection -
XX
XX
XX Disclosure; Page 302-309; 577pp; English.
XX
XX The present invention describes proteins and their fragments (I) encoded
XX
XX by chromosome 2 of the human malarial parasite, Plasmodium falciparum.
XX
XX Also described are: (i) nucleotide sequences (ii) encoding (I); and (2)
XX
XX vaccines against P. falciparum infection comprising (I) or (II).
XX
XX CC (I) and (II) are useful for the development of vaccines against
XX
XX P. falciparum infection. (I) and polyclonal antisera or a monoclonal
XX
XX CC antibody raised to immunogens comprising the sequences of (I), are
XX
XX useful in the detection of infection with P. falciparum. Furthermore,
XX
XX CC (I) (especially when they are rifins or secreted or membrane proteins)
XX
XX CC can aid the identification of drugs to treat or prevent P. falciparum
XX
XX CC infection, or they can be used to identify drug resistance in
XX
XX CC P. falciparum. Sequencing of the Plasmodium chromosome 2 and the
```



DR WPI; 2003-167481/16.  
DR N-PSDB; AB222903.  
XX  
PT New isolated antibody that binds to a staphylococcal surface protein,  
XX useful for treating or preventing Staphylococcus aureus infections,  
PT such as pneumonia, septic arthritis, endocarditis or biomaterial  
PT related infections -  
XX  
PS Claim 1; Page 29-30; 122pp; English.  
XX  
XX The present invention describes an isolated antibody (I) that binds to a  
CC staphylococcal surface protein selected from any of the 12 sequences of  
CC 354-2283 amino acids given in ABP56875 to ABP56886. Also described:  
CC (1) an isolated antisera (II) comprising (1); (2) a diagnostic kit (III)  
CC comprising (1) and means for detecting binding by (1); (3) diagnosing  
CC (M1) an infection of Staphylococcus aureus by adding (1) to a sample  
CC suspected of being infected with the infection, and determining if  
CC antibodies have bound to the sample; (4) a pharmaceutical composition  
CC (IV) for treating or preventing an infection of S. aureus comprising (1),  
CC and a vehicle, carrier or excipient; (5) treating (M2) or preventing an  
CC infection of S. aureus by administering (1) to a human or animal patient;  
CC (6) producing (M3) an immunological response by administering to a human  
CC or animal an immunogenic amount of the isolated surface protein; (7) an  
CC isolated active fragment (V) from the A domain of the DsgA protein; and  
CC (8) a vaccine (VI) for treating or preventing an infection of S. aureus  
CC comprising the surface protein in an amount effective to elicit an immune  
CC response, and a vehicle, carrier or excipient. (I) has antiinflammatory,  
CC antibacterial, immunosuppressive and antiarthritic activities, and can be  
CC used in gene therapy. The antibody, composition and vaccine are useful  
CC for treating or preventing Staphylococcus aureus infection in a human or  
CC animal, such as pneumonia, septic arthritis, endocarditis or biomaterial  
CC related infections. The present sequence represents Staphylococcus  
CC epidermidis KKM protein, which is used in the exemplification of  
CC the present invention.  
XX  
XX Sequence 654 AA:  
SQ  
Query Match 11.0%; Score 93.5; DB 24; Length 654;  
Best Local Similarity 21.9%; Pred. No. 2.2; Mismatches 74; Indels 67; Gaps 11;  
Matches 49; Conservative 34; Mismatches 74; Indels 67; Gaps 11;  
OY 1 TTVKEPILNKDGE-VSELKPRVTV--TIQNGKEMSTIVSEEDILPVYKGELE-KGY 56  
DB 336 SAITEFQNVQPTNEKKTDLQDTKYVYVESVENNESMDTFVGH-----PIKGMANGKXY 390  
OY 57 QF-----DGEWISGFEGRK-----D 71  
DB 391 WMEETTTNDYMDQFVWEGQVRTISDKAKNTRTIIFFVYEGKTLVDALVKVHVKTIDVD 450  
OY 72 AGYVINTL-SKDTFIPVFKKIBKKEEKKPTPDV-----SKKQNPQVNHGQLESNHRK 125  
DB 451 GQYHVRIVDKFAFTANTDKSNKKQSDNSAKKEATPATPSKPTSPVEKESQKODSORD 510  
OY 126 EDLQ-----REHSQKSDSTFDVT-ATVLDKNNISSKSTNNPNK 164  
DB 511 DNKQLPVSEKENDASSEGKDTPATKPTXGVESSSTT--PTK 552  
RESULT 21  
ID ABJ18979 standard; Protein; 645 AA.  
XX  
XX ABJ18979;  
XX  
XX 06-MAR-2003 (first entry)  
XX  
XX Pathogen specific antigen related staphylococcal protein SEQ ID No 142.  
XX  
XX Antibacterial; virucide; fungicide; protozoacide; cytostatic; anti-HIV;  
XX hyperimmune; serum-reactive; antigen; pathogen; tumour; allergen;  
XX auto-immunity; vaccine; staphylococcal infection; antibody; cancer;  
XX autoimmune disease; HIV; hepatitis.  
XX

OS Staphylococcus sp.  
XX  
XX PN W0200259148-A2.  
XX  
XX 01-AUG-2002.  
XX  
XX 21-JAN-2002; 2002WO-EP00546.  
XX  
XX 26-JAN-2001; 2001AT-0000130.  
XX  
XX (CIST-) CISTEM BIOTECHNOLOGIES GMBH.  
XX  
XX Meinke A, Nagy E, Von Ahnen U, Klade C, Henics T, Zauner W;  
XX Minh DB, Vytvytska O, Etz H, Dryla A, Weichart T, Hafner M;  
XX Tempelmeier B;  
XX WPI; 2003-075410/07.  
XX  
XX Identifying, isolating and producing hyperimmune serum-reactive  
PT antigens from a pathogen, for preparing vaccine or medicament for  
PT treating or preventing e.g. staphylococcal infections, comprises  
XX providing antibody preparation -  
XX  
PS Claim 21; Page 168; 252pp; English.  
XX  
XX The invention relates to a novel method for identifying, isolating and  
CC producing hyperimmune serum-reactive antigens from a pathogen, tumour,  
CC allergen, a tissue or host prone to auto-immunity, where the antigens  
CC are used in a vaccine, comprises providing antibody preparation from a  
CC plasma pool of a type of animal, or individual sera with antibodies  
CC against the specific pathogen, tumour, allergen, tissue or host prone to  
CC auto-immunity. The hyperimmune serum-reactive antigens comprising any of  
CC the 62 sequences of 53-2261 amino acids fully defined in the  
CC specification, or their hyperimmune fragments are useful for the  
CC manufacture of a pharmaceutical preparation, particularly a vaccine  
CC against staphylococcal infections or colonisation against S. aureus or S.  
CC epidermidis. The preparation of antibodies is useful for the manufacture  
CC of a medicament for treating or preventing staphylococcal infections or  
CC colonisation against S. aureus or S. epidermidis. The antibody  
CC preparations may also be used for diagnostic and imaging purposes. Other  
CC conditions that can be treated include cancer, autoimmune diseases or  
CC infections caused by viral (e.g. HIV, hepatitis A, B or C), fungal or  
CC protozoan pathogens. This sequence represents a staphylococcal protein  
CC relating to the method for identifying and producing pathogen specific  
CC antigens of the invention.  
XX  
XX Sequence 645 AA:  
SQ  
Query Match 10.9%; Score 92.5; DB 24; Length 645;  
Best Local Similarity 21.9%; Pred. No. 2.7; Mismatches 74; Indels 67; Gaps 11;  
Matches 49; Conservative 34; Mismatches 74; Indels 67; Gaps 11;  
OY 1 TTVKEPILNKDGE-VSELKPRVTV--TIQNGKEMSTIVSEEDILPVYKGELE-KGY 56  
DB 327 SAITEFQNVQPTNEKKTDLQDTKYVYVESVENNESMDTFVGH-----PIKGMANGKXY 381  
OY 57 QF-----DGEWISGFEGRK-----D 71  
DB 382 WMEETTTNDYMDQFVWEGQVRTISDKAKNTRTIIFFVYEGKTLVDALVKVHVKTIDVD 441  
OY 72 AGYVINTL-SKDTFIPVFKKIBKKEEKKPTPDV-----SKKQNPQVNHGQLESNHRK 125  
DB 442 GQYHVRIVDKFAFTANTDKSNKKQSDNSAKKEATPATPSKPTSPVEKESQKODSORD 501  
OY 126 EDLQ-----REHSQKSDSTFDVT-ATVLDKNNISSKSTNNPNK 164  
DB 502 DNKQLPVSEKENDASSEGKDTPATKPTXGVESSSTT--PTK 543  
RESULT 22  
ID AAG82397 standard; Protein; 299 AA.  
XX  
XX AAG82397



AC	AA02397;
XX	
DT	03-SEP-2001 (first entry)
DE	
XX	
XX	5. epidermidis open reading frame protein sequence SEQ ID NO:1888.
XX	
KW	Staphylococcus epidermidis SRI strain; infection; diagnosis;
XX	vaccination; endocarditis.
XX	
OS	Staphylococcus epidermidis.
XX	
PN	WO200134809-A2.
PD	
XX	17-MAY-2001.
XX	
PF	09-NOV-2000; 2000MO-US30782.
XX	
PR	09-NOV-1999; 99US-0164258.
XX	
PA	(GLAXO ) GLAXO GROUP LTD.
XX	
P1	Kimmerly WJ;
XX	
DR	WPI; 2001-316495/33.
XX	
DR	N-PSDB; AAH53247.
PT	
XX	Nucleic acids encoding polypeptides from Staphylococcus epidermidis,
XX	useful for vaccinating against infections, e.g. endocarditis -
PS	Claim 18; Page 516; 2188bp; English.
XX	
CC	AAH52304 to AAH53970 represent nucleic acids (i) encoding polypeptides
CC	(ii), given in AAH54154 to AAH53120, from Staphylococcus epidermidis.
CC	(i) and (ii) can have antibacterial activity and therefore can be used
CC	in vaccination. The nucleic acids (i) may be used to produce the
CC	S. epidermidis polypeptides (ii) via the production of vectors
CC	containing them which are used to produce hosts cells which express the
CC	polypeptides. The polypeptides (ii) (and/or nucleic acids) may then be
CC	used to vaccinate subjects and to raise antibodies against the bacteria.
CC	The polypeptides may also be used to assay for other inhibitors of their
CC	activity and therefore identify compounds that may be used for the
CC	treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to
CC	AAH5090 represent specifically claimed S. epidermidis genomic DNA
CC	polynucleotide sequences from the present invention. AAH5091 to
CC	AAH5098 represent oligonucleotide sequences and primers which are used
CC	in the exemplification of the present invention.
CC	N.B. The present invention specifically claims all the polynucleotide
CC	sequences given in the sequence listing of the present specification,
CC	however the sequence listing only goes up to SEQ ID NO:4454 so even
CC	though sequences are given in the disclosure for SEQ ID NO:4465 to 4472,
CC	no sequences are present for SEQ ID NO:4455 to 4464.
XX	
Sequence	299 AA;
Query Match	10.8%; Score 92; DB 22; Length 299;
Best Local Similarity	24.4%; Pred. No. 1;
Matches	39; Conservative 27; Mismatches 60; Indels 34; Gaps 5.
9	NDGTGEVBLKPHRTVTITQNGKMSSTVSEDPF-----LPVYAGLEKG-YQPDGWEI 63
33	NKDT-EKSDKXTHRIISLIPSTBELIYRIGIEDIVGVSTVDYFKDVKGGKQPDANL 91
64	SGFE-----GKDAGVIMLSKDTFKPV-----FKKIEBK 95
92	NKEHLIKAKPDILAHBSQKSNAGVLSLKDKGKRVVYVDKAGSIDETDFKSIQGLT 151
96	EEENKPTFVSKKQNDPQVNHSQLNESHKEDLQREHSQ 135
152	DREKQAKELVDETGNVEKINSVPEKHKKGVFVNEVSK 191

[illegible]

DE Staphylococcus epidermidis ORF amino acid sequence SEQ ID NO:3033.  
XX  
KM Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;  
XX antibacterial; gene therapy.  
XX  
OS Staphylococcus epidermidis.  
XX  
PN US6380370-B1.  
XX  
PD 30-APR-2002.  
XX  
PF 13-AUG-1998; 98US-0134001.  
XX  
PR 14-AUG-1997; 97US-055779P.  
XX 08-NOV-1997; 97US-064964P.  
XX  
PA (GENO-) GENOME THERAPEUTICS CORP.  
XX  
PI Doucette-Stamm LA, Bush D;  
XX  
DR WPI; 2002-381255/41.  
XX N-PSDB; ABN90733.  
XX  
PT Novel isolated nucleic acid encoding a Staphylococcus epidermidis  
XX polypeptide, useful for diagnosing and treating bacterial infections -  
XX  
PS Disclosure; SEQ ID 3033; 267bp; English.  
XX  
CC ABN90538 to ABN93374 represent Staphylococcus epidermidis open reading  
CC frame (ORF) nucleic acid sequences which encode the amino acid sequences  
CC given in ABP35124 to ABP37960. The S. epidermidis sequences have  
CC antibacterial activity and can be used in gene therapy. The sequences  
CC can also be used in the diagnosis and treatment of bacterial infections,  
CC particularly S. epidermidis infections. The sequences can be used to  
CC screen for compounds able to interfere with the S. epidermidis life  
CC cycle or inhibit S. epidermidis infection.  
CC N.B. The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from the  
CC USPTO web site.  
XX  
SQ Sequence 442 AA;  
  
Query Match 10.7%; Score 91; DB 23; Length 442;  
Best Local Similarity 21.7%; Pred. No. 2.2;  
Matches 49; Conservative 30; Mismatches 65; Indels 82; Gaps 11;  
  
QY 10 KDTGEVSELKPHRYVT-IONGE--MSSTVSEDPILPYKG-ELER----- 54  
DB 90 EDT-EINPVAQPEVNTQIEKGDFIFRATVTVPEVKLGDKYKLEIKQETDLSDEIQ 148  
QY 55 -----GYQPDGWEISG--FEGKQAGYVINISKDTPTK 85  
DB 149 ESIDSLSLAEVNVKEDAVENGDTVINIDFSG-SVDGSEFDGQAGYDLRIGSGSTFP 207  
QY 86 PVFKKLEBKKEEENK-----PTFDVS---KKQNPQVNHSQLNE 121  
DB 208 GFEEQIEGKKTGDKDVVTTPBRYAELAGKATFKKVAEIKFKQVPELANDELANEL 267  
QY 122 -----SHRKEDLQREHSHOKSDSTKDYATATVLDKNISKSTTN 160  
DB 268 DSDANVDEYKENVLRKRLSEQKATEAENT-----EKEBAINKATEN 308  
  
RESULT 25  
ABP40312  
ID ABP40312 standard; Protein; 902 AA.  
XX  
AC ABP40312;  
XX  
DT 24-JUL-2002 (first entry)  
XX  
DE Staphylococcus epidermidis ORF amino acid sequence SEQ ID NO:5157.  
XX

KM Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;  
XX antibacterial; gene therapy.  
XX  
OS Staphylococcus epidermidis.  
XX  
PN US6380370-B1.  
XX  
PD 30-APR-2002.  
XX  
PF 13-AUG-1998; 98US-0134001.  
XX  
PR 14-AUG-1997; 97US-055779P.  
XX 08-NOV-1997; 97US-064964P.  
XX  
PA (GENO-) GENOME THERAPEUTICS CORP.  
XX  
PI Doucette-Stamm LA, Bush D;  
XX  
DR WPI; 2002-381255/41.  
XX N-PSDB; ABN92857.  
XX  
PT Novel isolated nucleic acid encoding a Staphylococcus epidermidis  
XX polypeptide, useful for diagnosing and treating bacterial infections -  
XX  
PS Disclosure; SEQ ID 5157; 267bp; English.  
XX  
CC ABN90538 to ABN93374 represent Staphylococcus epidermidis open reading  
CC frame (ORF) nucleic acid sequences which encode the amino acid sequences  
CC given in ABP35124 to ABP37960. The S. epidermidis sequences have  
CC antibacterial activity and can be used in gene therapy. The sequences  
CC can also be used in the diagnosis and treatment of bacterial infections,  
CC particularly S. epidermidis infections. The sequences can be used to  
CC screen for compounds able to interfere with the S. epidermidis life  
CC cycle or inhibit S. epidermidis infection.  
CC N.B. The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from the  
CC USPTO web site.  
XX  
SQ Sequence 902 AA;  
  
Query Match 10.7%; Score 91; DB 23; Length 902;  
Best Local Similarity 24.9%; Pred. No. 6.1;  
Matches 43; Conservative 31; Mismatches 75; Indels 24; Gaps 8;  
  
QY 1 TTYKEPILNDGSEVSELKPHRYVTIONGKES-STVSEDPILPYKGLEKGYQPD 59  
DB 736 TPVSEYRLSNRGKGIK---TATITERNGNVICITTVGEEQLAVYTNAGV---IRLD 788  
QY 60 GWEISGFEGSKDAGY-VINISKDTFIKPVFKKLEBKKEEN-----KPTFDVSKK 109  
DB 789 VHDISQ-NGRAAQGVRLMTLGGQGFVSTYAKVNEEDNENMDEAQQSTTETADYBEV 847  
QY 110 DNPQVNHSQLNESHRKEDLQREB-HSOKSDSTKDYATATVLDKNISKSTTN 161  
DB 848 D-----DQTGNAIHTEGDALMESVESPENDDRIDIQDFMDRVNEDIESASDN 896  
  
RESULT 26  
AAB42897  
ID AAB42897 standard; Protein; 209 AA.  
XX  
AC AAB42897;  
XX  
DT 08-FEB-2001 (first entry)  
XX  
DE Human ORFX ORF2661 polypeptide sequence SEQ ID NO:5322.  
XX  
XX  
XX Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;  
XX vulnery; antiparasitic; antiparkinsonian; nootropic; neuroprotective;  
XX anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;  
XX immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;  
XX hypotensive; dermatological; immunosuppressive; antineoplastic;  
XX antiviral; antibacterial; antifungal; antineumatic; antihypoid;  
XX

KW anti-naemic; gene therapy; cancer; proliferative disorder; hypertension;  
 KW neurodegenerative disorder; osteoarthritis; graft vs host disease;  
 KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;  
 KW cholesterol ester storage; systemic lupus erythematosus; infection;  
 KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;  
 KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;  
 KW bone damage; cartilage damage; anti-inflammatory disease; coagulation;  
 KW thrombosis; contraceptive.  
 OS Homo sapiens.  
 XX  
 XX WO200058473-A2.  
 PN  
 XX  
 XX 05-OCT-2000.  
 PD  
 XX  
 XX 31-MAR-2000; 2000WO-US08621.  
 PF  
 XX 31-MAR-1999; 99US-0127607.  
 PR 02-APR-1999; 99US-0127636.  
 PR 05-APR-1999; 99US-0127728.  
 PR 30-MAR-2000; 2000US-0540763.  
 XX  
 XX (CURA-) CURAGEN CORP.  
 PA  
 XX  
 XX Shinkets RA, Leach M;  
 PI  
 XX WPI; 2000-602362/57.  
 DR N-PSDB; AAC77106.  
 DR  
 XX  
 XX Novel nucleic acids and peptides derived from open reading frame X,  
 PT useful for treating e.g. cancers, proliferative disorders,  
 PT neurodegenerative disorders and cardiovascular disease -  
 PT  
 XX  
 XX Claim 11; Page 4497-4498; 5507P; English.  
 XX  
 XX AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,  
 CC which represent the human ORF open reading frames 1 to 3161. The ORF  
 CC sequences have activities such as: cytostatic; hepatotropic; vlnetary;  
 CC antiproliferative; antiparkinsonian; nocrotropic; neuroprotective;  
 CC osteoplastic; anticonvulsant; antiaeritic; immunosuppressant;  
 CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;  
 CC antidiabetic; hypotensive; dermatological; immunosuppressive;  
 CC antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;  
 CC antichryoid; and anti-naemic. The sequences can be used for determining  
 CC the presence of or predisposition to, or preventing or treating  
 CC pathological conditions associated with an ORF-associated disorder. The  
 CC nucleic acids can be used to express ORF proteins in gene therapy  
 CC vectors. The proteins and nucleic acids may be used to treat cancers,  
 CC proliferative disorders, neurodegenerative disorders, osteoarthritis,  
 CC graft vs host disease, cardiovascular disease, diabetes mellitus,  
 CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus  
 CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,  
 CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,  
 CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,  
 CC nocturnal haemoglobinuria, anti-inflammatory disease; to enhance  
 CC coagulation; to inhibit thrombosis; and as a contraceptive.  
 CC  
 XX  
 XX Sequence 209 AA;  
 SQ  
 Query Match 10.7%; Score 90.5; DB 21; Length 209;  
 Best local Similarity 28.1%; Pred. No. 0.89;  
 Matches 47; Conservative 17; Mismatches 56; Indels 47; Gaps 9;  
 QY 13 GEVSEILKPHRYVTYIIONKESSTIVSE--EDFLPYKGLKGYOF---DGMW---62  
 DB 18 GBLOQSGSQL-----HGKSDSPVNYTEKKEITALRELTBLBRKLFTEQORSULMERLY 72  
 QY 63 -----ISGFBCKDAG-----YVNLSDTFPKVPFKTBKKBKBNKTFPVSKKD 110  
 DB 73 VEAKDQNGKGTGDKKKGSGSHRAKKSKKTFPGSV-----KETPFAMNNT 120  
 QY 111 NPQVNH-----SOLNESHKEDLOREHHSQSD--STKDYATATYLDK 150

DB 121 KEFVRHHEKIKQAKA-VKENLKKFSDSVSTFRRHPTOTKNIFDE 166  
 RESULT 27  
 ABP73209  
 ID ABP73209 standard; Protein; 635 AA.  
 XX  
 XX ABP73209;  
 AC  
 XX  
 XX 30-JAN-2003 (first entry)  
 DT  
 XX  
 XX  
 DE Candida albicans essential protein SEQ ID NO 7046.  
 XX  
 XX  
 XX Fungus; yeast; tetracycline; promoter; GRACE strain; biosynthesis;  
 KW signal transduction; DNA replication; cell division; growth;  
 KW proliferation; Candida albicans; fungicide; antifungal.  
 XX  
 XX Candida albicans.  
 OS  
 XX  
 XX WO200253728-A2.  
 PN  
 XX  
 XX 11-JUL-2002.  
 PD  
 XX  
 XX 26-DEC-2001; 2001WO-US49486.  
 PF  
 XX  
 XX 29-DEC-2000; 2000US-259128P.  
 PR 20-FEB-2001; 2001US-0792024.  
 PR 22-AUG-2001; 2001US-314050P.  
 XX  
 XX (ELIT-) ELITRA PHARM INC.  
 PA  
 XX  
 XX Roemer T, Jiang B, Boone C, Bussey H, Ohlsen KL;  
 PI  
 XX WPI; 2002-566694/60.  
 DR N-PSDB; ABZ31759.  
 DR  
 XX  
 XX Constructing strains for identifying gene products as effective targets  
 PT for therapeutic intervention, by inactivating in the strain one allele  
 PT of a gene and placing other allele of the gene under conditional  
 PT expression -  
 PT  
 XX  
 XX Claim 44; SEQ ID NO 7046; 167pp + Sequence Listing; English.  
 PS  
 XX The invention relates to constructing (M1) a strain of diploid fungal  
 CC cells in which both alleles of a gene are modified, comprising modifying  
 CC one allele by insertion or replacement by a cassette having an  
 CC expressible selectable marker and modifying other allele by  
 CC recombination, of a promoter replacement fragment with a heterologous  
 CC promoter, so that expression of the second allele is regulated by the  
 CC promoter. (M1) is useful for constructing a strain of diploid fungal  
 CC cells in which both alleles of a gene are modified. The diploid fungal  
 CC cells having both alleles modified are useful for identifying a gene that  
 CC is essential to the survival or growth of a fungus, a gene that  
 CC contributes to the virulence and/or pathogenicity of a fungus, a gene  
 CC that contributes to the resistance and/or pathogenicity of a fungus, a  
 CC agent, an antifungal agent that inhibits the growth of a diploid fungus  
 CC and for identifying a therapeutic agent for treatment of a mammalian  
 CC disease. (M1) is useful for identifying a compound which modulates the  
 CC activity of a gene product, preferably enzymatic activity, carbon  
 CC compound catabolism, biosynthetic, transporter, transcriptional,  
 CC translational, signal transduction, DNA replication and cell division  
 CC activity. The method is useful for identifying a compound having the  
 CC ability to inhibit growth or proliferation of C. albicans cells and for  
 CC treating infection by C. albicans. The present sequence is that of an  
 CC essential Candida albicans protein used in the method of the invention.  
 CC Note: The sequence data for this patent is not represented in the printed  
 CC specification but is based on sequence information supplied to Derwent by  
 CC the European Patent Office.  
 CC  
 XX  
 XX Sequence 635 AA;  
 SQ  
 Query Match 10.7%; Score 90.5; DB 23; Length 635;  
 Best local Similarity 22.2%; Pred. No. 4.2;

	Matches	38;	Conservative	40;	Mismatches	70;	Indels	23;	Gaps	8;
Qy	6	FILNKDTGEVSELXPH---	RTVTATLIONGKEM-----	SSITVSEEDFILPYKGLKELGY	56					
Db	64	YTAKE--KMGKFPSPWTRMPTFP	PKQIETLNMWVKSKELEBOEF---	EBDGTTPPY	117					
Qy	57	QPDGM---EISGF--EGKDAGYV	INLSK-DTPIKPVFKKI	BEKKEEENKPTPDVSNOK	109					
Db	118	FFEKYVQGLBIPRFVEDETVDL-	TWIETKEYSKKEERARERKER	ERKERERKRDLEBKCK	176					
Qy	110	DNPDVNHSQLNSHRKEDLQRE	EHNSQKSDSTKVDYATVLD	KNINSSKSTTN	160					
Db	177	QOQOQOQSQOQPNQIQIKDEK	QKQOTRNNITDKDSQNSDEP	PTVEAKKETD	227					

RESULT 28  
AAY44364  
ID AAY44364 standard; protein; 757 AA.

AC	AAV44364;
XX	
DT	14-MAR-2000 (first entry)

Human cell cycle regulation protein-5.

KM CCRP-5; cell cycle regulation protein-5; cell proliferation;  
KM cell proliferative disease; cancer; atherosclerosis; cirrhosis;  
KM hepatitis; psoriasis; immune system disorder; allergy; asthma;  
KM acquired immune deficiency syndrome; Crohn's disease; Blast method;  
KM rheumatoid arthritis; gene therapy; chromosomal mapping.

**Homo sapiens.**

PH	Key	Location/Qualifiers
PH	Key	Location/Qualifiers

FT	Modified-site	63	/note= "Potential glycosylation site"
----	---------------	----	---------------------------------------

FT	/note=	"Potential glycosylation site"
FT	479	/note=
Modified-site		"Potential glycosylation site"

	Modified-site	Potential glycosylation site"
FT	498	
PT		
PM		
Modifed-site	31	

FT	Modified-site	/note="Potential phosphorylation site"
FT	43	

FT	Modified-site	81	/note=	"Potential phosphorylation site"
FT				

Protein	Site	Phosphorylation
FT	Modified-site	/note= "Potential phosphorylation site"
FT	99	

PT	Modified-site	130	"/note="	"Potential	phosphorylation site
PT					

	Potential phosphorylation site
Modified-site PT	158
PT	/note=
FT	
EM	163

FT	/note= "Potential phosphorylation site
FT	163
Modified-site	

FT	Modified-site	184	/note=	"Potential phosphorylation site"
PT				

FT	/note=	Potential phosphorylation site
FT	Modified-site	188

FT	/note=
Modified-site	potential phosphorylation site
271	

	Modified-site	PT	F1	F1
Modified-site	287	287	287	287
PT	287	287	287	287
F1	287	287	287	287
F1	287	287	287	287

PT	/note= "Potential phosphorylation site
PT	289
Modified-site	

FT		/note= "Potential phosphorylation site"
FT	Modified-site	323 /note= "Potential phosphorylation site"
FT	Modified-site	343 /note= "Potential phosphorylation site"
FT	Modified-site	367 /note= "Potential phosphorylation site"
FT	Modified-site	370 /note= "Potential phosphorylation site"
FT	Modified-site	432 /note= "Potential phosphorylation site"
FT	Modified-site	488 /note= "Potential phosphorylation site"
FT	Modified-site	499 /note= "Potential phosphorylation site"
FT	Modified-site	207 /note= "Potential phosphorylation site"
FT	Modified-site	277 /note= "Potential phosphorylation site"
FT	Modified-site	298 /note= "Potential phosphorylation site"
FT	Modified-site	370 /note= "Potential phosphorylation site"
FT	Modified-site	419 /note= "Potential phosphorylation site"
FT	Modified-site	474 /note= "Potential phosphorylation site"
FT	Modified-site	488 /note= "Potential phosphorylation site"
FT	Modified-site	499 /note= "Potential phosphorylation site"
FT	Modified-site	528 /note= "Potential phosphorylation site"
FT	Modified-site	532 /note= "Potential phosphorylation site"
FT	Modified-site	539 /note= "Potential phosphorylation site"
FT	Modified-site	609 /note= "Potential phosphorylation site"
FT	Modified-site	624 /note= "Potential phosphorylation site"
FT	Modified-site	730 /note= "Potential phosphorylation site"
PX		
PN	W09964593-A2.	
PD	16-DEC-1999.	
PX		
PP	08-JUN-1999;	99WO-US12906.
PX	08-JUN-1998;	98US-0088695.
PA	(INCY-) INCYTE PHARM INC.	
PX		
PI	Bandman O, Lal P, Tang YT, Corley NC, Guegler KJ, Baughn MR; Patterson C;	
DR	WP1; 2000-105887/09.	
N-PSDB;	AAMZ29484.	
XK		
XX	Novel regulatory proteins, for diagnosis, treatment and prevention of cell proliferative and immune system diseases -	
PS	Claim 1; Page 70-71; 87pp; English.	
CC	The present sequence is cell cycle regulation protein-5 (CECRP-5). Blast analytical method was used to identify this protein. CECRPs are activators of cell proliferation or inhibitors of cellular processes that modulate proliferation. They are used to treat or prevent cell proliferative diseases like cancers, atherosclerosis, cirrhosis, hepatitis, psoriasis, immune system disorders (e.g. acquired immune deficiency syndrome, allergy, asthma, Crohn's disease, rheumatoid arthritis). Antibodies are raised to screen for specific binding agents.	

CC The corresponding nucleic acid is used in gene therapy, chromosomal  
CC mapping and isolation of related sequences.

XX  
SQ Sequence 757 AA;

Query Match 10.7%; Score 90.5; DB 21; Length 757;

Best Local Similarity 28.1%; Pred. No. 5.3;

Matches 47; Conservative 17; Mismatches 56; Indels 47; Gaps 9;

QY 13 GEVSELKPHRYTVITONGKEMSTIVSE--BDPILPVYKGLKGYQF-----DQWE--- 62

DB 398 GELQQLSGSQL-----HGKSDSPVYTEKEIALIRELTLEKRLTFEQGRSDLMWELLY 452

QY 63 -----ISGFEKGKQAG-----YVINLSKDTPIKVPYFKLIBKKEENKPTFVSKKD 110

DB 453 VEAKDQNGKQGTGDKKKGGSGSHRAKNKSKTFPGSV-----KETFDAMKNST 500

QY 111 NPOVNH-----SQINESHRKEDLQREHSQKSD--STQDVATATVLDK 150

DB 501 KEFVRHHEKIKQAKKA-VKENLKKFSDSVKSTFRHFDITTKNIFDE 546

#### RESULT 29

AAB53319 standard; Protein: 758 AA.

AC AAB53319;

DT 09-MAR-2001 (first entry)

XX Human colon cancer antigen protein sequence SEQ ID NO:859.

XX Human; colon cancer; colon cancer antigen; diagnosis; detection;

KM identification; cytostatic; cardioactive; neuroprotective; vulnary;

KM immunomodulatory; muscular; gynaecological; gastrointestinal;

KM nephrotropic; antiinfective; antibacterial; gene therapy; wound;

KM neural disorder; immune system disorder; muscular disorder;

KM reproductive disorder; gastrointestinal disorder; renal disorder;

XX infectious disease; cardiovascular disorder.

OS Homo sapiens.

PN MO200055351-A1.

XX 21-SEP-2000.

XX 08-MAR-2000; 2000MO-US05883.

XX 12-MAR-1999; 99US-0124270.

XX (HUMA-) HUMAN GENOME SCI INC.

PI Rosen CA, Ruben SM;

XX MPI: 2000-587534/55.

DR N-PSDB; AAC98076.

XX Colon cancer associated gene sequences, referred to as colon cancer

PT antigens, useful for the treatment, prevention, and diagnosis of colon

PT disorders such as colon cancer -

XX Claim 11, Page 1408-1411; 2104pp; English.

XX AAC97991 to AAC98763 encode the human colon cancer associated proteins,

CC called human colon cancer antigens, given in AAB53234 to AAB54006. The

CC human colon cancer antigens can have cytostatic, cardioactive, muscular;

CC neuroprotective, immunomodulatory, gynaecological, gastrointestinal,

CC vulnary, nephrotropic, antiinfective and antibacterial activities, and

CC can be used in gene therapy. The colon cancer antigen polynucleotides,

CC proteins and antibodies to the proteins are useful for the prevention,

CC treatment and diagnosis of colon disorders, such as colon cancer. The

CC polynucleotides may be used in diagnostics and research, such as for

CC chromosome identification, and as hybridisation probes. The proteins

CC may also be used to prevent diseases such as neural disorders, immune

CC system disorders, muscular disorders, reproductive disorders,

CC gastrointestinal disorders, wounds, renal disorders, infectious

CC diseases, and cardiovascular disorders. AAC98764 to AAC98772 and

CC AAB54007 represent sequences used in the exemplification of the present

CC invention.

XX Sequence 758 AA;

Query Match 10.7%; Score 90.5; DB 21; Length 758;

Best Local Similarity 28.1%; Pred. No. 5.3;

Matches 47; Conservative 17; Mismatches 56; Indels 47; Gaps 9;

QY 13 GEVSELKPHRYTVITONGKEMSTIVSE--BDPILPVYKGLKGYQF-----DQWE--- 62

DB 399 GELQQLSGSQL-----HGKSDSPVYTEKEIALIRELTLEKRLTFEQGRSDLMWELLY 453

QY 63 -----ISGFEKGKQAG-----YVINLSKDTPIKVPYFKLIBKKEENKPTFVSKKD 110

DB 454 VEAKDQNGKQGTGDKKKGGSGSHRAKNKSKTFPGSV-----KETFDAMKNST 501

QY 111 NPOVNH-----SQINESHRKEDLQREHSQKSD--STQDVATATVLDK 150

DB 502 KEFVRHHEKIKQAKKA-VKENLKKFSDSVKSTFRHFDITTKNIFDE 547

#### RESULT 30

AAB68963 standard; Protein: 801 AA.

AC AAB68963;

DT 20-JAN-2003 (first entry)

XX Human polypeptide SEQ ID NO 1010.

XX Human; genome mapping; gene therapy; food supplement; virus; fungus;

KM cell-proliferative disorder; neurodegenerative disease; bacterial;

KM Parkinson's disease; Alzheimer's disease; autoimmune disease;

KM multiple sclerosis; diabetes; genetic disorder; wound; burn; infection;

KM arthritis; cytostatic; immunomodulator; neotropic; dermatological;

KM antiparkinsonian; antidiabetic; immunosuppressive; dermatological;

KM haemostatic; vulnary; fungicide; antibacterial; virucide; protozoacide;

XX antiarthritic.

OS Homo sapiens.

PN WO200270539-A2.

XX 12-SEP-2002.

XX 05-MAR-2002; 2002MO-US05095.

XX 05-MAR-2001; 2001US-0799451.

XX (HYSE-) HYSEQ INC.

PI Tang YT, Zhou P, Goodrich RW, Asundi V, Zhang J, Zhao QA, Ren F;

PI Xue AJ, Yang Y, Ma Y, Yamazaki V, Chen R, Wang Z, Ghosh M;

PI Wehrman T, Wang J, Wang D, Drmanac RT;

XX MPI: 2002-759812/82.

DR N-PSDB; ABZ11180.

XX New polynucleotides comprising sequences assembled from expressed

PT sequence tags (ESTs), useful for creating cell-proliferative,

PT neurodegenerative, autoimmune, genetic, myeloid or lymphoid, or

PT platelet or coagulation disorders -

XX Claim 9; SEQ ID NO 1010; 1012pp + Sequence Listing; English.

XX The invention relates to an isolated polynucleotide (I) comprising a

CC nucleotide sequence selected from any of 948 sequences



useful for diagnosing diseases related to over or underexpression of *tig* protein by identifying mutations in the *tig* gene, or determining *tig* polypeptide or mRNA expression levels due to an infection of an organism with the *tig* gene. They can diagnose the stage and type of infection. *Tig* polypeptides are also useful for screening compounds which affect activity of the protein by measuring the binding to *tig* and observing the stimulation or inhibition of the polypeptide function. These can be used in treatment to inhibit or enhance *tig* activity, in addition to direct administration of *tig* polypeptides to treat conditions associated with a lack of *tig* polypeptide, or direct administration of antisense sequences to prevent expression. *Tig* polypeptides (administered directly, in a vector and as a vaccine) and antibodies induce an immune response to immunise and prevent disease. Diseases diagnosed, prevented or treated include: bacterial infections, especially *Staphylococcus aureus* infections which cause otitis media, toxic shock syndrome, conjunctivitis, wound infection, impetigo and septic arthritis etc. *Tig* polypeptides, polynucleotides and their (ant)agonists can prevent adhesion of bacteria to matrix proteins, and are useful for use on wounds and body implants to prevent bacterial infection.

XX Genome sequence of Chlamydia pneumoniae  
PT Page 975-976; Disclosure; 1912pp; English.  
XX  
XX AAY34584-Y35879 represent the proteins encoded by all the open reading  
CC frames in the complete genome (see AAY91990) of Chlamydia pneumoniae.  
CC C. pneumoniae causes respiratory disease such as pneumonia and  
CC bronchitis and is thought to be a contributing factor in heart  
CC disease, sarcoidosis, sinusitis, purulent otitis media, erythema  
CC nodosum or pharyngitis. The polypeptides encoded by the open reading  
CC frames of the C. pneumoniae genome (see AAY34584-Y35879) can be used in  
CC immunogenic compositions as vaccines. Vectors containing C. pneumoniae  
CC nucleotide sequences can also be used as immunogenic compositions,  
CC especially where the vector directs the expression of a neutralising  
CC epitope of C. pneumoniae.  
XX  
SQ Sequence 511 AA;  
Query Match 10.6%; Score 89.5; DB 20; Length 511;  
Best Local Similarity 24.5%; Pred. No. 3.9;  
Matches 23; Conservative 17; Mismatches 33; Indels 21; Gaps 2;  
QY 84 IKPVFKKIKBEKKBEKKPTFD-----VSKKKNPQVNHSQLNBSHRK 125  
DB 95 VKGVFKKTPQARPRVSSPRLPSHVHGQRLPLGLEGRDRIQKRSNPADLGKKRSYSYD 154  
QY 126 EDLQREHSHQSKSTQVATATVDKNNISKSTT 159  
DB 155 GDLDRVGHDSNBDSTEDSRS---EGGEPSKSSS 185  
RESULT 35  
ABP73570 standard; Protein; 516 AA.  
ID ABP73570 standard; Protein; 516 AA.  
XX  
XX ABP73570;  
XX  
XX 30-JAN-2003 (first entry)  
XX  
XX Candida albicans essential protein SEQ ID NO 7407.  
XX  
XX  
XX Fungus; yeast; tetracycline; promoter; GRACE strain; biosynthesis;  
XX signal transduction; DNA replication; cell division; growth;  
XX proliferation; Candida albicans; fungicide; antifungal.  
XX  
XX  
XX Candida albicans.  
XX  
XX WO200253728-A2.  
XX  
XX  
XX 11-JUL-2002.  
XX  
XX  
XX 26-DEC-2001; 2001WO-US49486.  
XX  
XX  
XX 29-DEC-2000; 2000US-259128P.  
XX 20-FEB-2001; 2001US-0792024.  
XX 22-AUG-2001; 2001US-314050P.  
XX  
XX  
XX (ELIT-) ELITRA PHARM INC.  
XX  
XX  
XX Roemer T, Jiang B, Boone C, Bussey H, Ohlsen KL;  
XX  
XX WPI; 2002-566694/60.  
XX N-PSDB; AB232120.  
XX  
XX  
XX Constructing strains for identifying gene products as effective targets  
XX for therapeutic intervention, by inactivating in the strain one allele  
XX of a gene and placing other allele of the gene under conditional  
XX expression -  
XX  
XX Claim 44; SEQ ID NO 7407; 167pp + Sequence Listing; English.  
XX  
XX The invention relates to constructing (M1) a strain of diploid fungal

CC cells in which both alleles of a gene are modified, comprising modifying  
CC one allele by insertion or replacement by a cassette having an  
CC expressible selectable marker and modifying other allele by  
CC recombination, of a promoter replacement fragment with a heterologous  
CC promoter, so that expression of the second allele is regulated by the  
CC promoter. (M1) is useful for constructing a strain of diploid fungal  
CC cells in which both alleles of a gene are modified. The diploid fungal  
CC cells having both alleles modified are useful for identifying a gene that  
CC is essential to the survival or growth of a fungus, a gene that  
CC contributes to the virulence and/or pathogenicity of a fungus, a gene  
CC that contributes to the resistance and/or pathogenicity of a fungus, a  
CC agent, an antifungal agent that inhibits the growth of a diploid fungus  
CC and for identifying a therapeutic agent for treatment of a mammalian  
CC disease. (M1) is useful for identifying a compound which modulates the  
CC activity of a gene product, preferably enzymatic activity, carbon  
CC compound catabolism, biosynthetic, transporter, transcriptional,  
CC translational, signal transduction, DNA replication and cell division  
CC activity. The method is useful for identifying a compound having the  
CC ability to inhibit growth or proliferation of C. albicans cells and for  
CC treating infection by C. albicans. The present sequence is that of an  
CC essential Candida albicans protein used in the method of the invention.  
CC Note: The sequence data for this patent is not represented in the printed  
CC specification but is based on sequence information supplied to Derwent by  
CC the European Patent Office.  
XX  
SQ Sequence 516 AA;  
Query Match 10.6%; Score 89.5; DB 23; Length 516;  
Best Local Similarity 23.3%; Pred. No. 3.9;  
Matches 37; Conservative 27; Mismatches 58; Indels 37; Gaps 6;  
QY 10 KONGEVELKPRHYVT--IQNGKMSSTVSEDEFLPYRKGLENGYGP----- 58  
DB 369 KNGKISRYLANCKSIASRIDNYSBPTTAFGB-----ILKKQVEDRLFPYDGSAPMK 422  
QY 59 -----DGMELSGFGKKDAGVYNLSKDTFKPKVKKIBKKK---EENKPTF 103  
DB 423 NSDAITKALALNQGDLAAGABQKQVDMV---SDEBQVKKKKKKKKKKKKKKKKKK 479  
QY 104 DVSKKKNPQVNHSQLNBSHRKEDLQREHSHQSKSTKD 142  
DB 480 DKKEKDKKKKKKK--DKKRSDDGEETPKKKKKKKSKD 516  
RESULT 36  
AAY03189 standard; Protein; 525 AA.  
ID AAY03189 standard; Protein; 525 AA.  
XX  
XX AAY03189;  
XX  
XX  
XX 16-JUN-1999 (first entry)  
XX  
XX  
XX S. aureus trigger factor protein sequence.  
XX  
XX  
XX Trigger factor; rig; diagnosis; immune response; bacterial infection;  
XX S. aureus infection; otitis media; toxic shock syndrome; conjunctivitis;  
XX wound infection; impetigo; septic arthritis; therapy.  
XX  
XX  
XX Staphylococcus aureus.  
XX  
XX  
XX  
XX EP905233-A2.  
XX  
XX  
XX 31-MAR-1999.  
XX  
XX  
XX 21-AUG-1998; 98EP-0306697.  
XX  
XX  
XX 29-DEC-1997; 97US-0999339.  
XX 04-SEP-1997; 97US-0057511.  
XX  
XX  
XX (SMK ) SMITHKLINE BEECHAM CORP.  
XX (SMK ) SMITHKLINE BEECHAM PLC.  
XX  
XX Burnham MKR, Fosberry A, Hodgson JE, Jaworski DD;



PI Lawlor EJ, Rosenberg M, Traini CM, Wang M, Ward J;  
PI Warren RL;  
XX  
XX WPI; 1999-192663/17.  
DR N-PSDB; AAX28174.  
XX  
XX New Staphylococcus aureus trigger factor (tig) polypeptide and  
PT polynucleotide, useful as diagnostic reagents and for prevention and  
PT treatment of Staphylococci aureus infections  
XX  
XX

PS Claim 14; Page 6; 31pp; English.

XX This sequence is the Staphylococcus aureus trigger factor (tig)  
CC polypeptide of the invention. Tig polypeptides and polynucleotides are  
CC useful for diagnosing diseases related to over or underexpression of tig  
CC protein by identifying mutations in the tig gene, or determining tig  
CC polypeptide or mRNA expression levels due to an infection of an organism  
CC with the tig gene. They can diagnose the stage and type of infection. Tig  
CC polypeptides are also useful for screening compounds which affect  
CC activity of the protein by measuring the binding to tig and observing the  
CC stimulation or inhibition of the polypeptide function. These can be used  
CC in treatment to inhibit or enhance tig activity, in addition to direct  
CC administration of tig polypeptides to treat conditions associated with a  
CC lack of tig polypeptide, or direct administration of antisense sequences  
CC to prevent expression. Tig polypeptides (administered directly, in a  
CC vector and as a vaccine) and antibodies induce an immune response to  
CC immunise and prevent disease. Diseases diagnosed, prevented or treated  
CC include: bacterial infections, especially Staphylococcus aureus  
CC infections which cause otitis media, toxic shock syndrome,  
CC conjunctivitis, wound infection, impetigo and septic arthritis etc. Tig  
CC polypeptides, polynucleotides and their (ant)agonists can prevent  
CC adhesion of bacteria to matrix proteins, and are useful for use on wounds  
CC and body implants to prevent bacterial infection.  
XX

SQ Sequence 525 AA;

Query Match 10.6%; Score 89.5; DB 20; Length 525;  
Best Local Similarity 22.1%; Pred. No. 4;  
Matches 50; Conservative 33; Mismatches 60; Indels 83; Gaps 13;

QY 11 DTGVSLSLKHRRVVT-IQNGKE--MSTIVSEDFILPYVKG-ELER----- 54  
DB 81 DDTDIKVAQPEVSTOIEKGDPIFEATVVEPEVKLGDKGLEIEKQETELSDDELQ 140  
QY 55 -----GYQPDGWEISG--FRKKDAGVIMLSKOTFIKP 86  
DB 141 AIDHSLGLHLMVYKEDGVENGDTVINDFSG-SVDGEFEGQAGEGDLSIGSGSFI-P 198  
QY 87 VFK-----KIEBKKE-----EE--NKPTPDVS---KKKDNPOVNHSQLNE- 121  
DB 199 GFEEQLGEMKVDKDVVVTPEEYHAEELAGKATFTKYNELKFEVPELITDEIANEL 258  
QY 122 -----SHRKEDLQREHESQKSDSTKDVATVLDKNNISKSTTN 160  
DB 259 DABANTVDEYKENARKRLAEQKATDAENV-----EKERATIKATDN 299

RESULT 37  
ABB58704  
ID ABB58704 standard; Protein; 1408 AA.

XX ABB58704;  
XX  
XX  
XX 26-MAR-2002 (first entry)

DE Drosophila melanogaster polypeptide SEQ ID NO 2904.

XX Drosophila; developmental biology; cell signalling; insecticide;  
XX pharmaceutical.  
XX

OS Drosophila melanogaster.  
XX  
XX

PN WO200171042-A2.

XX  
PD 27-SEP-2001.  
XX  
XX 23-MAR-2001; 2001WO-US09231.  
XX  
XX  
XX 23-MAR-2000; 2000US-191637P.  
PR 11-JUL-2000; 2000US-0614150.  
XX  
XX (PEKE ) PE CORP NY.  
XX  
XX  
XX Venter JC, Adams M, Li FWD, Myers EW;  
XX  
XX WPI; 2001-656860/75.  
DR N-PSDB; ABL02807.

XX New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions -  
XX  
XX  
XX Disclosure; SEQ ID NO 2904; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABL01840-ABL16175), expressed DNA  
CC sequences (ABL57737-ABR72072).  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pcc\_sequences.  
XX

SQ Sequence 1408 AA;

Query Match 10.6%; Score 89.5; DB 22; Length 1408;  
Best Local Similarity 24.7%; Pred. No. 16;  
Matches 48; Conservative 23; Mismatches 62; Indels 61; Gaps 9;

QY 29 NGKE-----MSTIVSEDFP--ILPYKGELEKGYQFDGWEISGFGK--KDAQVY----- 75  
DB 646 NGKQLNASKETKTSEKDFKVEVSKKE-BKNDQESKLENDQKLPKDDDYLDVVKP 704  
QY 76 -----INLSKDTPIKPVFK-----KIEKKEEENKPTFDVSKKKDNPQVNHSQLN 120  
DB 705 MESKESYSEPLAVKLGEDVSKSPSMSSSDTKISVVGGE--PRMNGBLPKTGSQVQRPKKS 762  
QY 121 ESHRK-EDLQREHESQKSDSTKDVY-----ATVLDK 150  
DB 763 KTOAKIEPLAVVQSELKSKSVKDYVTKEQGTTKBETHKGFSPKREATPKSESEKGTTEVDK 822  
QY 151 NNISKSTTNPNPK 164  
DB 823 SNDVSVIKRSPDP 836

RESULT 38  
ABB93341  
ID ABB93341 standard; Protein; 296 AA.

XX ABB93341;  
XX

XX 31-MAY-2002 (first entry)

DE Herbicidally active polypeptide SEQ ID NO 2552.

XX Herbicidal; plant; agriculture; herbicide.

XX Arabidopsis thaliana.  
XX

OS WO200210210-A2.  
XX  
XX

PD 07-FEB-2002.

```

XX 28-AUG-2001; 2001WO-EP09892.
XX
XX 28-AUG-2001; 2001WO-EP09892.
XX
XX (FARB ) BAYER AG.
XX
XX Tietjen K, Weidler M;
XX
XX WPI; 2002-269010/31.
XX
XX Identifying plant target proteins for herbicidally active compounds,
XX comprising aligning and comparing nucleic acid or amino acid sequences
XX from plant with nucleic acid or amino acid sequences from non-plant
XX organisms -
XX
XX Claim 5; SEQ ID NO 2552; 261bp + Sequence Listing; English.
XX
XX The invention relates to identifying target proteins
XX (ABB90790-ABB94016) for herbicidally active compounds, comprising
XX aligning and comparing nucleic acid or amino acid sequences from plant
XX with nucleic acid or amino acid sequences from non-plant organisms using
XX suitable search parameters, where plant sequences having an E-value
XX greater by a factor of 3 than the E-value of most similar non-plant
XX sequences are selected. The polypeptides or nucleic acids encoding them
XX are useful for identifying modulators. The identified modulators are
XX useful as herbicides.
XX
XX Sequence 296 AA;
XX
XX Query Match 10.5%; Score 89; DB 23; Length 296;
XX Best Local Similarity 23.1%; Pred. No. 2;
XX Matches 43; Conservative 33; Mismatches 70; Indels 40; Gaps 7;
XX
XX 3 VKFPIINKOTGEVSELEKPRVTVTTONGK-----EMSTIVSEDFILP---- 46
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX 26 ISQVALGESKNNTB--PIQLVTVGSDKLLIGTLSEKFPOLSTEIVLERNFALSHTWK 83
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX 47 ----VKGSELEKGYQPDGMEISGFEKGKAGYININLSKDTFIPVKIKBEKKEENKPT 102
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX 84 NGSVPFSGYKPEDLIDDLAAGAFKGMGLY-----PPAAPKSAKQVNPQLPN 133
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX 103 FVYSKKKDN---POVNHSQLNSHREKEDLQREHSOKSDSTQDVATVLDKNINISKST 158
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX 134 EDVKAKQDDADGSESDSDSDSENSGDEBEKVTASDSERDSD--DESDSSEER 191
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX
XX 159 TNNPNK 164
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX 192 T--PKK 195
XX
XX RESULT 39
XX ABB64828
XX ID ABB64828 standard; Protein; 281 AA.
XX
XX ABB64828;
XX
XX 26-MAR-2002 (first entry)
XX
XX Drosophila melanogaster polypeptide SEQ ID NO 21276.
XX
XX Drosophila; developmental biology; cell signalling; insecticide;
XX pharmaceutical.
XX
XX Drosophila melanogaster.
XX
XX WO200171042-A2.
XX
XX 27-SEP-2001.
XX
XX 23-MAR-2001; 2001WO-US09231.
XX
XX 23-MAR-2000; 2000US-191637P.
XX

```

```

PR 11-JUL-2000; 2000US-0614150.
XX
XX (PEKE ) PE CORP NY.
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
XX
XX WPI; 2001-656860/75.
XX
XX N-PSDB; ABL08931.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from Drosophila and for elucidating cell signalling and cell-cell
XX interactions -
XX
XX Disclosure; SEQ ID NO 21276; 21bp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signalling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (ABL16176-ABL10511), expressed DNA
XX sequences (ABL01840-ABL16175) and the encoded proteins
XX (ABB5737-ABB72072).
XX
XX The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 281 AA;
XX
XX Query Match 10.4%; Score 88.5; DB 22; Length 281;
XX Best Local Similarity 22.1%; Pred. No. 2.1;
XX Matches 31; Conservative 34; Mismatches 52; Indels 23; Gaps 5;
XX
XX 29 NGKEMSTIVSEDFILPVYKGELEKGYQPDGMEISGFEKGKAGYI-----NLS 79
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX 137 NFEEFTNMTTQDDLKNIT-SLDDK-----DNEGATISKELGVIRALGRPNESIA 188
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX 80 KDTFIPVFPKIRE-KKEENKPTPDVSKKQDNPOVNHSQLNE-----SHRKEDLQREH 133
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX 189 KESEFCVILNKKMDTNEBELRDAFRVFDKNNGYISTELRAVFMALGERLEDEDEEM 248
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX
XX 134 SQRSDSTQDVATVLDKNNI 153
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX 249 IREYDLDQDNHNFESNNM 268
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX
XX RESULT 40
XX AAG85008
XX ID AAG85008 standard; Protein; 1141 AA.
XX
XX AAG85008;
XX
XX 11-SEP-2001 (first entry)
XX
XX Shrimp white spot Bacilliform virus (WSBV) protein 99.
XX
XX Shrimp white spot Bacilliform virus; WSBV; diagnosis; viral infection;
XX antiviral agent; gene expression; antisense construct;
XX transgenic viral resistant shrimp.
XX
XX white spot syndrome virus.
XX
XX WO200138351-A2.
XX
XX 31-MAY-2001.
XX
XX 08-NOV-2000; 2000WO-US28888.
XX
XX 24-NOV-1999; 99CN-0124717.
XX
XX (PENY-) PE CORP NY.
XX
XX (THIR-) THIRD INST OCEANOGRAPHY STATE OCEANI C A.
XX
XX (SINO-) SINOGENOMAX CO LTD.
XX

```

